arXiv:1112.5634v1 [math.ST] 23 Dec 2011

MODEL SELECTION FOR POISSON PROCESSES WITH COVARIATES

MATHIEU SART

Abstract. We observe n inhomogeneous Poisson processes with covariates and aim at estimating their intensities. To handle this problem, we assume that the intensity of each Poisson process is of the form $s(\cdot, x)$ where x is the covariate and where s is an unknown function. We propose a model selection approach where the models are used to approximate the multivariate function s. We show that our estimator satisfies an oracle-type inequality under very weak assumptions both on the intensities and the models. By using an Hellinger-type loss, we establish non-asymptotic risk bounds and specify them under various kind of assumptions on the target function s such as being smooth or composite. Besides, we show that our estimation procedure is robust with respect to these assumptions.

1. Introduction

We consider n independent Poisson point processes N_i for $i = 1, \dots, n$ with values in the measurable space $(\mathbb{T}, \mathscr{T})$. For each i, we assume that the intensity of N_i with respect to some reference measure μ on $(\mathbb{T}, \mathscr{T})$ is of the form $s_i(\cdot) = s(\cdot, x_i)$ where x_i is a deterministic element of some measurable set $(\mathbb{X}, \mathscr{X})$ and s is a non-negative function on $\mathbb{T} \times \mathbb{X}$ satisfying

$$\forall i \in \{1, \dots, n\}, \quad \int_{\mathbb{T}} s(t, x_i) \, \mathrm{d}\mu(t) < +\infty.$$

Typically, this corresponds to the modelling of the times of failure of n repairable systems where the reliability of each of them depends on external factors measured by some covariates $x_1, \dots, x_n$, in which case $\mathbb{T}$ corresponds to an interval of time, say $[0, 1]$, and $\mathbb{X}$ to some compact subset of $\mathbb{R}^k$, say $[0, 1]^k$. Our aim is to estimate s from the observations of the pairs $(N_i, x_i)_{1 \leq i \leq n}$.

Let $\mathbb{L}^1_+(\mathbb{T} \times \mathbb{X}, M)$ be the cone of integrable and non negative functions on $(\mathbb{T} \times \mathbb{X}, \mathscr{T} \otimes \mathscr{X})$ equipped with the product measure $M = \mu \otimes \nu_n$ where $\nu_n = n^{-1} \sum_{i=1}^n \delta_{x_i}$. In order to evaluate the risks of our estimators, we endow $\mathbb{L}^1_+(\mathbb{T} \times \mathbb{X}, M)$ with the

Date: December, 2011.

2000 *Mathematics Subject Classification.* 62G05, 62M30.

Key words and phrases. Adaptive estimation, Intensity estimation, Model selection, Poisson processes with covariates, T-estimator.

Hellinger-type distance H defined for $u,v \in \mathbb{L}^1_+(\mathbb{T}\times\mathbb{X},M)$ by

$$\begin{aligned} 2H^2(u,v) &= \int_{\mathbb{T}\times\mathbb{X}} \left(\sqrt{u(t,x)}-\sqrt{v(t,x)}\right)^2 \mathrm{d}M(t,x) \\ &= \frac{1}{n}\sum_{i=1}^n \int_{\mathbb{T}} \left(\sqrt{u(t,x_i)}-\sqrt{v(t,x_i)}\right)^2 \mathrm{d}\mu(t). \end{aligned}$$

Let $(\mathbb{L}^2(\mathbb{T}\times\mathbb{X},M),d_2)$ be the metric space of functions f on $\mathbb{T}\times\mathbb{X}$ such that f^2 belongs to $\mathbb{L}^1_+(\mathbb{T}\times\mathbb{X},M)$. Given a suitable collection $\mathbb{V}$ of models (not necessarily linear spaces) and a non-negative application Δ on $\mathbb{V}$ satisfying

$$\sum_{V\in\mathbb{V}} e^{-\Delta(V)} \leq 1,$$

we build an estimator $\hat{s}$ whose risk $\mathbb{E}\left[H^2(s,\hat{s})\right]$ satisfies

$$C\mathbb{E}\left[H^2(s,\hat{s})\right] \leq \inf_{V\in\mathbb{V}}\left\{d_2^2\left(\sqrt{s},V\right)+\eta_V^2+\frac{\Delta(V)}{n}\right\}, \tag{1}$$

where C is an universal positive contant, $d_2\left(\sqrt{s},V\right)$ is the $\mathbb{L}^2$-distance between $\sqrt{s}$ and V and $n\eta_V^2$ is the metric dimension (in a suitable sense) of V. We shall use this inequality in order to derive risk bound for our estimator under smoothness or structural assumptions on the target function s.

In the literature many attention have been paid to the problem of estimating the intensity of a Poisson process without covariates. With the $\mathbb{L}^2$-loss, Reynaud-Bouret (2003) dealt with the problem of model selection among a family of linear spaces V. Baraud and Birgé (2009) used the Hellinger distance and considered the case where the sets V consist of piecewise constants functions on a partition of $\mathbb{T}$. More general models were considered by Birgé (2007) allowing for V any subsets with finite metric dimensions (in a suitable sense). Much less is known when the intensity s depends on covariates. For model selection purpose, the only result we are aware of is due to Comte et al. (2011) where they considered the $\mathbb{L}^2$-loss and penalized projection estimators on linear spaces. Their approach requires that the intensity s be bounded from above by a quantity that needs to be either known or suitably estimated. Besides, they impose some restrictions on the family of linear spaces V so that their estimator possesses minimax properties over classes of functions which are smooth enough.

Our approach is based on robust testing as developed in the papers of Birgé (2006) and Baraud (2010). We shall see that our estimator $\hat{s}$ possesses nice (adaptation and robustness) properties but suffers from the fact that its construction is numerically intractable. From this point of view, it inherits from the qualities and drawbacks of the T-estimators as developed by Birgé (2006). We obtain an oracle inequality of the form (1) under very mild assumptions both on the intensity s and the family of models $\mathbb{V}$. This allows us to derive risk bounds over a large range of Hölderian spaces including irregular ones. Nevertheless, we shall also consider functions s defined on a subset $\mathbb{T}\times\mathbb{X}$ of a large dimensional linear space, say $\mathbb{T}\times\mathbb{X}=[0,1]^{1+k}$

with k large, and it is well known that in such a situation, the minimax approach based on smoothness assumptions may lead to very slow rates of convergence. This phenomenon is known as *the curse of dimensionality*. In this case, an alternative approach is to assume that s belongs to classes $\mathscr{F}$ of functions satisfying structural assumptions (such as the multiple index model, the generalized additive model, the multiplicative model ...) and for which faster rates of convergence can be achieved. Very recently, this approach was developed by Juditsky et al. (2009) (in the Gaussian white noise model) and by Baraud and Birgé (2011) (in more statistical settings). Nevertheless, unlike Juditsky et al. (2009) we shall not assume that s belongs to $\mathscr{F}$ but rather consider $\mathscr{F}$ as an approximating class for s.

In the present paper, our point of view is closer to that developed in Baraud and Birgé (2011). We shall use our new model selection theorem in conjunction with suitable families $\mathbb{V}$ of models in order to design an estimator $\hat{s}$ possessing good statistical properties with respect to many classes of functions of interest, including classes $\mathscr{F} = \mathscr{F}_\circ$ consisting of composites functions $(t,x) \mapsto g(t, u(x))$ and classes $\mathscr{F} = \mathscr{F}_\times$ consisting of products functions $(t,x) \mapsto g(t)u(x)$. In order to design the suitable family $\mathbb{V}$ with good approximations properties with respect to the elements of $\mathscr{F}$, we shall either use the techniques developed in Baraud and Birgé (2011) when $\mathscr{F} = \mathscr{F}_\circ$ for instance, but also some new ones when $\mathscr{F} = \mathscr{F}_\times$ and for which the former approach by Baraud and Birgé would lead to an extra (and unnecessary) logarithmic factor in the risk bound. When $s(t,x)$ is of the form (or close to) $g(t,u(x))$ or $g(t)u(x)$, where g and u are assumed to be smooth we shall prove that our estimator is fully adaptive with respect to the regularities of both g and u. We shall also consider structural assumptions on the functions g and u as well as parametric ones when t and x lie in a large dimensional space. Finally we shall look at the situation where, for all $x \in \mathbb{X}$, $s(\cdot, x)$ belongs to a parametric class of functions $\mathscr{F}_\Theta$ with $\Theta \subset \mathbb{R}^k$, which means that there exists some element $f_{\theta(x)} \in \mathscr{F}_\Theta$ such that $s(\cdot, x) = f_{\theta(x)}(\cdot)$, and our aim is then to estimate the mapping $x \mapsto \theta(x)$ by model selection.

This paper is organized as follows. The general model selection theorem can be found in Section 2. In Section 3, we study the case where $\mathscr{F}$ is a class of smooth or composite functions, and in Section 4 the case where $\mathscr{F}$ is a class of product functions. The problem of estimating s when the intensity of each Poisson process N_i belongs to a same parametric model is dealt in Section 5. Section 6 is devoted to the proofs.

Let us introduce some notations that will be used all along the paper. We set $\mathbb{N}^\star = \mathbb{N} \setminus \{0\}$, $\mathbb{R}^\star = \mathbb{R} \setminus \{0\}$ and $\mathbb{R}_+^\star = \mathbb{R}_+ \cap \mathbb{R}^\star$. The numbers $x \wedge y$ and $x \vee y$ stand for $\min(x,y)$ and $\max(x,y)$ respectively. For $(E, \mathcal{E}, \nu)$ a measured space, we denote by $\mathbb{L}^2(E, \nu)$ the linear space of measurable functions f such that $\int_E |f|^2 \, \mathrm{d}\nu < \infty$. When $(E, \nu) = (\mathbb{T} \times \mathbb{X}, M)$, the $\mathbb{L}^2$-distance of this set is denoted by d_2, and the norm by $\|\cdot\|_2$. Alternatively, this distance (respectively this norm) is denoted by $d_{\mathtt{t}}$ (respectively $\|\cdot\|_{\mathtt{t}}$) when $(E, \nu) = (\mathbb{T}, \mu)$, and by $d_{\mathtt{x}}$ (respectively $\|\cdot\|_{\mathtt{x}}$) when $(E, \nu) = (\mathbb{X}, \nu_n)$. The supremum norm of a bounded function f on a domain E

is denoted by $\|f\|_\infty = \sup_{x\in E} |f(x)|$, and the space of all bounded functions on E by $\mathbb{L}^\infty(E)$. For (E,d) a metric space, $x \in E$ and $A \subset E$, the distance between x and A is denoted by $d(x,A) = \inf_{a\in A} d(x,a)$. The closed ball centered at $x \in E$ with radius r is denoted by $\mathcal{B}(x,r)$. The cardinality of a finite set A is denoted by $|A|$. The set $\mathscr{F}$ is a generic notation for a family of functions of $\mathbb{L}^2(\mathbb{T}\times\mathbb{X}, M)$ of special interest. The notations C,C',C''... are for the constants. The constants C,C',C''... may change from line to line.

2. A general model selection theorem

2.1. **Main result.** Throughout this paper, a model V is a subset of $\mathbb{L}^2(\mathbb{T}\times\mathbb{X}, M)$ with bounded metric dimension, in the sense of Definition 6 of Birgé (2006). We recall this definition below.

Definition 2.1. *Let V be a subset of $\mathbb{L}^2(\mathbb{T}\times\mathbb{X}, M)$ and D_V a right-continuous map from $(0,+\infty)$ into $[1/2,+\infty)$ such that $D_V(\eta) = o\big(\eta^2\big)$ when $\eta \to +\infty$. We say that V has a metric dimension bounded by D_V if for all $\eta > 0$, there exists $S_V(\eta) \subset \mathbb{L}^2(\mathbb{T}\times\mathbb{X}, M)$ such that for all $f \in \mathbb{L}^2(\mathbb{T}\times\mathbb{X}, M)$, there exists $g \in S_V(\eta)$ with $d_2(f,g) \le \eta$ and such that*

$$\forall \varphi \in \mathbb{L}^2(\mathbb{T}\times\mathbb{X}, M),\, \forall x \ge 2, \quad |S_V(\eta) \cap \mathcal{B}(\varphi, x\eta)| \le \exp\left(D_V(\eta)\, x^2\right).$$

Moreover, if one can choose D_V as a constant, we say that V has a finite metric dimension bounded by D_V.

This notion is more general than the dimension for linear spaces since a linear space V with finite dimension (in the usual sense) has a finite metric dimension. Besides, if V is not reduced to $\{0\}$ one can choose $D_V = \dim V$, what we shall do along this paper. Other models of interest with bounded metric dimension will appear later in the paper.

Given a collection of such subsets, our approach is based on model selection. We propose a selection rule based on robust testing in the spirit of the papers Birgé (2006); Baraud (2010). The test and the selection rule which are mainly abstract are postponed to Section 6. The main result is the following.

Theorem 2.1. *Let $\mathbb{V}$ be an at most countable family of models V with bounded metric dimension $D_V(\cdot)$ and Δ be some mapping from $\mathbb{V}$ into $\mathbb{R}_+$ such that $\sum_{V\in\mathbb{V}} e^{-\Delta(V)} \le 1$.*

One can build an estimator $\hat{s} \in \mathbb{L}^1_+(\mathbb{T}\times\mathbb{X}, M)$ depending on $(N_i, x_i)_{1\le i\le n}$, $\mathbb{V}$ and Δ such that

$$C\mathbb{E}\left[H^2(s,\hat{s})\right] \le \inf_{V\in\mathbb{V}} \left\{ d_2^2\left(\sqrt{s}, V\right) + \eta_V^2 + \frac{\Delta(V)}{n} \right\}, \tag{2}$$

where C is an universal positive constant and where

$$\eta_V = \inf\left\{\eta > 0,\ \frac{D_V(\eta)}{\eta^2} \le n\right\}.$$

Moreover, there exists a random function $\hat{f} \in \cup_{V\in\mathbb{V}} V$ such that $\sqrt{\hat{s}} = \hat{f} \vee 0$.

The condition $\sum_{V\in\mathbb{V}} e^{-\Delta(V)} \le 1$ can be interpreted as a (sub)probability on the collection $\mathbb{V}$. The more complex the family $\mathbb{V}$, the larger the weights $\Delta(V)$. When $\mathbb{V}$ consists of linear spaces V of finite dimensions D_V one can take $\eta_V^2 = D_V/n$ and hence (2) leads to

$$C\mathbb{E}\left[H^2(s,\hat{s})\right] \le \inf_{V\in\mathbb{V}} \left\{ d_2^2\left(\sqrt{s}, V\right) + \frac{D_V + \Delta(V)}{n} \right\}.$$

When one can choose $\Delta(V)$ of order D_V, which means that the family $\mathbb{V}$ of models does not contains too many models per dimension, the estimator $\hat{s}$ achieves the best trade-off (up to a constant) between the approximation and the variance terms.

In the remaining part of this paper, we shall consider subsets $\mathscr{F} \subset \mathbb{L}^2(\mathbb{T}\times\mathbb{X}, M)$ corresponding to various assumptions on $\sqrt{s}$ (smoothness, structural, parametric assumptions ...). For such an $\mathscr{F}$, we associate a collection $\mathbb{V}_{\mathscr{F}}$ and deduce from Theorem 2.1 a risk bound for the estimator $\hat{s}$ whenever $\sqrt{s}$ belongs or is close to $\mathscr{F}$. This bound takes the form

$$C'\mathbb{E}\left[H^2(s,\hat{s})\right] \le \inf_{f\in\mathscr{F}} \left\{ d_2^2(\sqrt{s}, f) + \varepsilon_{\mathscr{F}}(f) \right\} \tag{3}$$

where

$$\varepsilon_{\mathscr{F}}(f) = \inf_{V\in\mathbb{V}_{\mathscr{F}}} \left\{ d_2^2(f, V) + \eta_V^2 + \frac{\Delta(V)}{n} \right\},$$

and we shall bound by above the term $\varepsilon_{\mathscr{F}}(f)$. This upper bound will mainly depend on some properties of f, for example smoothness ones. In this case, this result says that if $\sqrt{s}$ is irregular but sufficiently close to a smooth function f, the bound we get essentially corresponds to the one we would get for f. This can be interpreted as a robustness property.

Sometimes, several assumptions on $\sqrt{s}$ are plausible, and one does not know what class $\mathscr{F}$ should be taken. A solution is to consider $\mathfrak{F}$ a collection of such classes $\mathscr{F}$ and to use the proposition below to get an estimator whose risk satisfies (up to a remaining term) relation (3) simultaneously for all classes $\mathscr{F} \in \mathfrak{F}$.

Proposition 2.1. *Let $\mathfrak{F}$ be an at most countable collection of subsets of $\mathbb{L}^2(\mathbb{T}\times\mathbb{X}, M)$ and Δ be some mapping on $\mathfrak{F}$ into $\mathbb{R}_+$ such that $\sum_{\mathscr{F}\in\mathfrak{F}} e^{-\Delta(\mathscr{F})} \le 1$. For any $\mathscr{F} \in \mathfrak{F}$, let $\mathbb{V}_{\mathscr{F}}$ be a collection of models and $\Delta_{\mathscr{F}}$ be some mapping such that the assumptions of Theorem 2.1 hold.*

There exists an estimator $\hat{s}$ such that, for all $\mathscr{F} \in \mathfrak{F}$,

$$C\mathbb{E}\left[H^2(s,\hat{s})\right] \le \inf_{f\in\mathscr{F}} \left\{ d_2^2(\sqrt{s}, f) + \varepsilon_{\mathscr{F}}(f) \right\} + \frac{\Delta(\mathscr{F})}{n},$$

where

$$\varepsilon_{\mathscr{F}}(f) = \inf_{V\in\mathbb{V}_{\mathscr{F}}} \left\{ d_2^2(f, V) + \eta_V^2 + \frac{\Delta_{\mathscr{F}}(V)}{n} \right\},$$

and where C is an universal positive constant.

We illustrate below Theorem 2.1 with two preliminary examples of $\mathscr{F}$. More general families $\mathscr{F}$ will be studied subsequently.

2.2. **Robustness with respect to an i.i.d assumption.** In this section, $\mathbb{X} = \{1, \dots, n\}$, which means that we observe n independent Poisson processes N_i, $1 \leq i \leq n$ with respective intensity $s_i(\cdot) = s(\cdot, i)$ on $\mathbb{T}$. Assuming that the N_i are i.i.d amounts to assuming that $\sqrt{s}$ belongs to

$$\mathscr{F} = \left\{ f \in \mathbb{L}^2(\mathbb{T} \times \mathbb{X}, M),\ f(\cdot, i) = f(\cdot, j),\ \forall 1 \leq i, j \leq n \right\},$$

which means that the dependency with respect to the second variable can be dropped. Given now a family of models $\mathbb{V}$ for approximating functions of one variable in $\mathbb{L}^2(\mathbb{T}, \mu)$, we deduce from Theorem 2.1 the following result.

Proposition 2.2. *Let $\mathbb{V}$ be an at most countable family of models V included into $\mathbb{L}^2(\mathbb{T}, \mu)$ with bounded metric dimension $D_V(\cdot)$ and Δ be some mapping from $\mathbb{V}$ into $\mathbb{R}_+$ such that $\sum_{V \in \mathbb{V}} e^{-\Delta(V)} \leq 1$.*

There exists an estimator $\hat{s} \in \mathbb{L}^1_+(\mathbb{T}, \mu)$ such that, for all $f \in \mathbb{L}^2(\mathbb{T}, \mu)$,

$$C\mathbb{E}\left[H^2(s, \hat{s})\right] \leq \frac{1}{n} \sum_{i=1}^{n} d_{\mathtt{t}}^2\left(\sqrt{s_i}, f\right) + \inf_{V \in \mathbb{V}} \left\{ d_{\mathtt{t}}^2(f, V) + \eta_V^2 + \frac{\Delta(V)}{n} \right\},$$

where C is an universal positive constant.

If the Poisson processes N_i were i.i.d, the preceding proposition could be proved from a model selection theorem for a single Poisson process by considering the aggregated process $\sum_{i=1}^{n} N_i$. We refer to the papers Baraud and Birgé (2009); Baraud (2010); Birgé (2007) for such theorems. However, our result still holds when the i.i.d assumption is slightly violated, since, in such a case, the risk of our estimator $\hat{s}$ cannot be severely deteriorated. This corresponds to some robustness with respect to this assumption.

Under some smoothness or structural assumptions on f, and for a suitable choice of $\mathbb{V}$, this bound can be specified in order to get rate of convergence in the same way as we shall do in Section 3.

2.3. **Time-dependent covariates.** In this section, we consider the case where we have at hand n repairable systems such that the times of failure of each of them can be modelled by a Poisson process. For each system, we assume that the probability of failure at time t depends only on t and the values of some measurements $(x(u))_{u \leq t}$ which have been recorded up to time t. This means that for each $i \in \{1, \dots n\}$, the intensity s_i of the Poisson process N_i representing the times of failure of system i, is of the form $s_i(t) = s\left(t, (x_i(u))_{u \leq t}\right)$. In this section, we therefore consider the case where $\mathbb{T} = \mathbb{R}_+$, and the covariate x is a function from $\mathbb{T}$ into $\mathbb{R}^k$ (and hence $\mathbb{X}$ is the set on functions from $\mathbb{R}_+$ into $\mathbb{R}^k$). A convenient simplification to model this kind of situation is to assume that $s_i(t)$ actually depends on t and $x_i(t)$ and not on the

past values $(x_i(u))_{u<t}$. Since the problem amounts then to estimating a function on $\mathbb{R}_+ \times \mathbb{R}^k$, we introduce the class

$$\mathscr{F} = \left\{ f \in \mathbb{L}^2(\mathbb{T} \times \mathbb{X}, M),\, \exists g \in \mathbb{L}^\infty(\mathbb{T} \times \mathbb{R}^k),\, \forall (t,x) \in \mathbb{T} \times \mathbb{X},\, f(t,x) = g(t, x(t)) \right\}.$$

The preceding simplification holds when $\sqrt{s}$ does belong to this class. We assume in this section that $\sqrt{s}$ is close to (but not necessarily an element of) $\mathscr{F}$ which corresponds to some robustness with respect to this simplification. We deduce from Theorem 2.1 the following result.

Proposition 2.3. *Assume that μ is a probability measure on $\mathbb{T}$. Let $\mathbb{V}$ be a family of finite dimensional linear subspaces of $\mathbb{L}^\infty(\mathbb{T} \times \mathbb{R}^k)$ and $\Delta \geq 1$ be some mapping on $\mathbb{V}$ such that $\sum_{V \in \mathbb{V}} e^{-\Delta(V)} \leq 1$. There exists an estimator $\hat{s}$, such that, for all $f \in \mathscr{F}$, and $g \in \mathbb{L}^\infty(\mathbb{T} \times \mathbb{R}^k)$ such that f is of the form $f(t,x) = g(t, x(t))$,*

$$C\mathbb{E}\left[H^2(s, \hat{s})\right] \leq d_2^2\left(\sqrt{s}, f\right) + \inf_{V \in \mathbb{V}} \left\{ d_\infty^2(g, V) + \frac{\dim V + \Delta(V)}{n} \right\}, \tag{4}$$

where C is an universal positive constant.

If $\sqrt{s}$ did belong to $\mathscr{F}$, the right-hand side of inequality (4) corresponds to the bound we would get if we could estimate a function g with $1+k$ variables by model selection. This bound can be specified under some assumptions on g in the same way as we shall do in the next section.

3. Some classical classes $\mathscr{F}$.

In this section, our aim is to control the quantity $\varepsilon_{\mathscr{F}}(f)$ appearing in (3) for various classes of interest $\mathscr{F}$. Throughout this section, we shall assume that μ is the Lebesgue measure.

3.1. **Classes of smooth functions.**

Let $\mathbf{I} = \prod_{j=1}^k I_j$ where the I_j are intervals of $\mathbb{R}$ and $\boldsymbol{\alpha} = \boldsymbol{\beta} + \mathbf{p} \in \left(\mathbb{R}_+^\star\right)^k$ with $\mathbf{p} \in \mathbb{N}^k$ and $\boldsymbol{\beta} \in]0,1]^k$. A function f belongs to the Hölder class $\mathcal{H}^{\boldsymbol{\alpha}}(\mathbf{I})$, if there exists $L(f) \in \mathbb{R}_+$ such that for all $(x_1, \ldots, x_k) \in \mathbf{I}$ and all $1 \leq j \leq k$, the functions $f_j(x) = f(x_1, \ldots, x_{j-1}, x, x_{j+1}, \ldots, x_k)$ admit a derivative of order $\mathbf{p}_j$ satisfying

$$\left| f_j^{(\mathbf{p}_j)}(x) - f_j^{(\mathbf{p}_j)}(y) \right| \leq L(f) |x - y|^{\boldsymbol{\beta}_j} \quad \forall x, y \in I_j.$$

The class $\mathcal{H}^{\boldsymbol{\alpha}}(\mathbf{I})$ is said to be isotropic when the α_j are all equal, and anisotropic otherwise, in which case $\bar{\boldsymbol{\alpha}}$ given by $\bar{\boldsymbol{\alpha}}^{-1} = k^{-1} \sum_{i=1}^k \alpha_i^{-1}$ corresponds to the average smoothness of a function f in $\mathcal{H}^{\boldsymbol{\alpha}}(\mathbf{I})$. We define the class of Hölderian functions on $\mathbf{I}$ by

$$\mathcal{H}(\mathbf{I}) = \bigcup_{\boldsymbol{\alpha} \in \left(\mathbb{R}_+^\star\right)^k} \mathcal{H}^{\boldsymbol{\alpha}}(\mathbf{I}).$$

Assuming that $\sqrt{s}$ is Hölderian corresponds thus to the choice $\mathscr{F} = \mathcal{H}(\mathbb{T}\times\mathbb{X})$. Anisotropic classes of smoothness are of particular interest in our context since the function s depends on variables t and x that may play very different roles.

Families of linear spaces possessing good approximation properties with respect to the elements of $\mathscr{F}$ can be found in the literature. We refer to the results of Dahmen et al. (1980). We may use these linear spaces (models) to approximate the elements of $\mathscr{F}$, and deduce from Theorem 2.1 the following result.

Corollary 3.1. *Let us assume that $\mathbb{T}\times\mathbb{X} = [0,1]^k$. There exists an estimator $\hat{s}$ such that, for all $f \in \mathcal{H}\left([0,1]^k\right)$, the inequality below holds:*

$$C\mathbb{E}\left[H^2(s,\hat{s})\right] \leq d_2^2(\sqrt{s}, f) + \max\left(L(f)^{\frac{2k}{2\bar{\alpha}+k}}\, n^{-\frac{2\bar{\alpha}}{2\bar{\alpha}+k}}, \frac{1}{n}\right) \tag{5}$$

where $\boldsymbol{\alpha} \in (\mathbb{R}_+^\star)^k$ is such that $f \in \mathcal{H}^{\boldsymbol{\alpha}}([0,1]^k)$ and where $C > 0$ depends only on k and $\max_{1\leq j\leq k}\alpha_j$.

Remark that the risk bound given by inequality (5) holds without any restriction on $\boldsymbol{\alpha}$. Such a generality can be obtained since our model selection theorem is valid for any collection $\mathbb{V}$ of finite dimensional linear spaces. Some restrictions on the dimensionality of the linear spaces $V \in \mathbb{V}$ (as in Comte et al. (2011)) would prevent us to get this rate of convergence for the Hölder classes $\mathcal{H}^{\boldsymbol{\alpha}}\left([0,1]^k\right)$ when $\min_{1\leq j\leq k}\alpha_j$ is too small.

The preceding risk bound is quite satisfactory if k is small but becomes worse when k increases. We shall therefore consider other types of classes in the next section in order to avoid this *curse of dimensionality.*

3.2. **Approximation by composite functions.** An alternative to the smoothness assumptions, is to consider structural ones. In this section we focus on the case where $\sqrt{s}$ is equal (or at least close to) a composite function of the form $g \circ u$ where g is a continuous mapping from $\mathbb{R}^l$ into $\mathbb{R}$ and u maps $\mathbb{T}\times\mathbb{X}$ into a compact subset of $\mathbb{R}^l$.

We shall restrict ourself to two cases of composite functions that are of particular interest for the problem we are considering. Nevertheless, more general composite functions $g \circ u$ could be handled in a similar way as in Baraud and Birgé (2011) leading to an analogue of their Theorem 1.

All along this section we shall use the following notation. For $g \in \mathcal{H}^{\boldsymbol{\alpha}}(\mathbf{I})$ where $\mathbf{I} = \prod_{j=1}^k I_j$, we denote by $\|g\|_{\boldsymbol{\alpha}}$ any number such that, for all $(x_1,\ldots,x_k) \in \mathbf{I}$, the function $g_j(x) = g(x_1,\ldots,x_{j-1},x,x_{j+1},\ldots,x_k)$ satisfies

$$\forall x, y \in I_j, \quad |g_j(x) - g_j(y)| \leq \|g\|_{\boldsymbol{\alpha}}|x-y|^{\alpha_j\wedge 1}.$$

3.2.1. *A multiple index model for covariate effect.* In this section, we fix $l \in \mathbb{N}^\star$ and we consider the case where $\sqrt{s}$ is equal (or close to) an element of

$$\begin{aligned}\mathscr{F}_l &= \Big\{ f \in \mathbb{L}^2(\mathbb{T}\times\mathbb{X}, M),\, \forall (t,x)\in\mathbb{T}\times\mathbb{X},\, f(t,x) = g\left(t, <\theta_1, x>, \ldots, <\theta_l, x>\right),\\ &\qquad g \in \mathcal{H}(\mathbb{T}\times[-1,1]^l),\, \forall 1\le \ell\le l,\ \theta_\ell \in \mathcal{B}(0,1)\Big\},\end{aligned}$$

where $\mathbb{T}=[0,1]^{k_1}$ and where $\mathbb{X}=\mathcal{B}(0,1)=\{x\in\mathbb{R}^{k_2},\, \sum_{i=1}^{k_2} x_i^2 \le 1\}$ is the unit ball of $\mathbb{R}^{k_2}$. This class $\mathscr{F}_l$ is related to the multiple index model and is introduced to reduce the curse of dimensionality when k_2 is large.

Corollary 3.2. *There exists an estimator $\hat{s}$ such that, for all $\boldsymbol{\alpha} \in (\mathbb{R}_+^\star)^{k_1+l}$, $\theta_1,\ldots,\theta_l \in \mathcal{B}(0,1)$, $g\in\mathcal{H}^{\boldsymbol{\alpha}}([0,1]^{k_1}\times[-1,1]^l)$ and $f\in\mathscr{F}_l$ of the form*

$$f(t,x) = g\left(t, <\theta_1, x>, \ldots, <\theta_l, x>\right),$$

the following inequality holds:

$$\begin{aligned} C\mathbb{E}\left[H^2(s,\hat{s})\right] &\le d_2^2(\sqrt{s}, f) + (L(g))^{\frac{2(k_1+l)}{2\bar{\alpha}+k_1+l}}\, n^{-\frac{2\bar{\alpha}}{2\bar{\alpha}+k_1+l}}\\ &\quad + \frac{\ln n \vee \left(\ln \|g\|_{\boldsymbol{\alpha}}^2 k_2^{-1}\right)\vee 1}{n} k_2 \end{aligned} \tag{6}$$

where $C>0$ depends only on l, k_1, and $\boldsymbol{\alpha}$.

In the case where $\sqrt{s}$ belongs to $\mathscr{F}_l$, $\sqrt{s}$ is an Hölderian function with k_1+k_2 variables and one could apply Corollary 3.1 to estimate the parameter s. However, it is easy to see that the preceding bound is much faster that the one we could get under smoothness assumption only. As soon as $l<k_2$, the larger term of (6) corresponds to the estimation rate of an Hölderian function g with only k_1+l variables and not k_1+k_2.

By using Proposition 2.1 with $\mathfrak{F}=\{\mathscr{F}_l,\, l\in\mathbb{N}^\star\}$, we derive an estimator $\hat{s}$ whose risk satisfies inequality (6) simultaneously for all $l\in\mathbb{N}^\star$.

3.2.2. *An Accelerated Failure-Time model.* The Accelerated Failure-Time model is a model in which the covariates change the time scale of some reference Poisson process N on $\mathbb{R}_+$. The Poisson process N typically corresponds of the times of failure of a repairable system in standard conditions. Here, we have at hand n systems, and we assume that for each $i\in\{1,\ldots,n\}$, and each $t\in\mathbb{R}_+$, the average number of failures of system i, before time t, $\mathbb{E}\left[N_i([0,t])\right]$ is of the form $\mathbb{E}\left[N_i([0,t])\right] = \mathbb{E}\left[N([0,tu(x_i)])\right]$ where u is a non-negative function on $\mathbb{X}$. This means that if $u(x_i)=1$, system i runs in normal condition and if $u(x_i)<1$ (respectively $u(x_i)>1$) the covariates extend (respectively reduce) the life-time of system i. Throughout this section, we assume that we observe each system $i\in\{1,\ldots,n\}$ on the interval of time $\mathbb{T}=[0,1]$. If the model is exact, and if g^2 denotes the underlying intensity of N, the above relation gives that s satisfies $\sqrt{s(t,x)} = v(x)g(tv^2(x))$, for all $(t,x)\in[0,1]\times\mathbb{X}$ where $v(x)=\sqrt{u(x)}$. As usual, we shall not assume that the model is exact, but rather that $\sqrt{s}$ is close to a function f of the form

$$f(t,x) = v(x)g(tv^2(x)) \quad \forall (t,x)\in[0,1]\times\mathbb{X}, \tag{7}$$

where g and v are smooth unknown functions. More precisely, we consider that $\mathbb{X} = [0,1]^{k_2}$ and we introduce the class of functions $\mathscr{F}$ given by

$$\begin{aligned}\mathscr{F} \quad = \quad & \Big\{ f \in \mathbb{L}^2([0,1] \times [0,1]^{k_2}, M),\, \exists v \in \mathcal{H}([0,1]^{k_2}),\, \exists g \in \mathcal{H}\left([0, \|v\|_\infty^2]\right), \\ & \forall (t,x) \in [0,1] \times [0,1]^{k_2},\, f(t,x) = v(x) g(tv^2(x)) \Big\}.\end{aligned}$$

The result is the following.

Corollary 3.3. *There exists an estimator $\hat{s}$ such that for all $f \in \mathscr{F}$ of the form (7) with $v \in \mathcal{H}^{\boldsymbol{\beta}}\left([0,1]^{k_2}\right)$, $v^2 \in \mathcal{H}^{\boldsymbol{\gamma}}\left([0,1]^{k_2}\right)$, $g \in \mathcal{H}^{\boldsymbol{\alpha}}\left([0, \|v\|_\infty^2]\right)$ for some $\alpha \in \mathbb{R}_+^\star$, $\boldsymbol{\beta} \in (\mathbb{R}_+^\star)^{k_2}$, $\boldsymbol{\gamma} \in (\mathbb{R}_+^\star)^{k_2}$, the following inequality holds:*

$$\begin{aligned} C\mathbb{E}\left[H^2(s, \hat{s})\right] \leq\ & d_2^2\left(\sqrt{s}, f\right) + \left(\|g\|_\infty L(v)\right)^{\frac{2k_2}{2\bar{\boldsymbol{\beta}} + k_2}} \left(\frac{\ln n \vee \ln\left(\|v\|_\infty \|g\|_\infty\right) \vee 1}{n}\right)^{\frac{2\bar{\boldsymbol{\beta}}}{2\bar{\boldsymbol{\beta}}+k_2}} \\ & + \left(\|v\|_\infty \|g\|_\alpha L\left(v^2\right)^{\alpha \wedge 1}\right)^{\frac{2k_2}{2\bar{\boldsymbol{\gamma}}(\alpha\wedge 1)+k_2}} \left(\frac{\ln n \vee \ln\left(\|v\|_\infty^{2(\alpha\wedge 1)+1} \|g\|_\alpha\right) \vee 1}{n}\right)^{\frac{2\bar{\boldsymbol{\gamma}}(\alpha\wedge 1)}{2\bar{\boldsymbol{\gamma}}(\alpha\wedge 1)+k_2}} \\ & + \left(\|v\|_\infty^{1+2\alpha} L(g)\right)^{\frac{2}{2\alpha+1}} n^{-\frac{2\alpha}{2\alpha+1}} + \frac{\ln n \vee \ln\left(\left(1 \vee \|v\|_\infty^{2(\alpha\wedge 1)}\right) \|v\|_\infty \|g\|_\alpha\right) \vee 1}{n} \end{aligned}$$

where $C > 0$ depends only on k_2, α, $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$.

Let us make some comments on the inequality above. If $\sqrt{s}$ does belong to $\mathscr{F}$ and if one is only interested by the rate of convergence, we can bound the risk of $\hat{s}$ by

$$C'\mathbb{E}\left[H^2(s,\hat{s})\right] \leq \max\left(\left(\frac{1}{n}\right)^{\frac{2\alpha}{2\alpha+1}}, \left(\frac{\ln n}{n}\right)^{\frac{2\bar{\boldsymbol{\beta}}}{2\bar{\boldsymbol{\beta}}+k_2}}, \left(\frac{\ln n}{n}\right)^{\frac{2\bar{\boldsymbol{\gamma}}(\alpha\wedge 1)}{2\bar{\boldsymbol{\gamma}}(\alpha\wedge 1)+k_2}}\right) \tag{8}$$

for some constant $C' > 0$ depending on s. The first term corresponds to the rate of convergence for estimating g only. If g is at least Lipschitz, the two other terms correspond to the rate of estimation of v and v^2 respectively (up to the logarithmic term). Note that it is always possible to choose $\boldsymbol{\gamma} = \boldsymbol{\beta}$, since v^2 is at least as regular as v. In which case, the rate becomes

$$C'\mathbb{E}\left[H^2(s,\hat{s})\right] \leq \max\left(\left(\frac{1}{n}\right)^{\frac{2\alpha}{2\alpha+1}}, \left(\frac{\ln n}{n}\right)^{\frac{2\bar{\boldsymbol{\beta}}(\alpha\wedge 1)}{2\bar{\boldsymbol{\beta}}(\alpha\wedge 1)+k_2}}\right).$$

Nevertheless, in some situations, v^2 may be more regular than v (think for instance to $v(x) = \sqrt{x}$ on $\mathbb{X} = [0,1]$) and hence, if $\bar{\boldsymbol{\gamma}}$ is large enough ($\bar{\boldsymbol{\gamma}} \geq \bar{\boldsymbol{\beta}}(\alpha \wedge 1)^{-1}$) the rate we get becomes

$$C'\mathbb{E}\left[H^2(s,\hat{s})\right] \leq \max\left(\left(\frac{1}{n}\right)^{\frac{2\alpha}{2\alpha+1}}, \left(\frac{\ln n}{n}\right)^{\frac{2\bar{\boldsymbol{\beta}}}{2\bar{\boldsymbol{\beta}}+k_2}}\right).$$

It is interesting to compare the rate (8) to the one we would obtain under the pure smoothness assumption on $\sqrt{s}$ but ignoring that $\sqrt{s}$ is of the form (7). To do so, we need to specify the regularity of $\sqrt{s}$, knowing that of v and g. This is the purpose of the following lemma. For the sake of simplicity $k_2 = 1$, $\alpha \in]0,1]$ and $\beta \in]0,1]$.

Lemma 3.1. *Let $\alpha \in]0,1]$, $\beta \in]0,1]$, $v \in \mathcal{H}^{\beta}([0,1])$ and $g \in \mathcal{H}^{\alpha}\left([0, \|v\|_{\infty}^2]\right)$. Then, the function f defined by $f(t,x) = v^2(x)g(tv(x))$ belongs to $\mathcal{H}^{(\alpha,\alpha\beta)}([0,1]^2)$.*

Moreover, there exists $v \in \mathcal{H}^{\beta}([0,1])$ and $g \in \mathcal{H}^{\alpha}\left([0, \|v\|_{\infty}^2]\right)$ such that, for all $\alpha' \in]0,1]$ and all $\beta' \in]0,1]$, the preceding function $f(t,x) = v^2(x)g(tv(x))$ belongs to $\mathcal{H}^{(\alpha',\alpha'\beta')}([0,1]^2)$ if and only if $\alpha' \leq \alpha$ and $\beta' \leq \beta$.

If $k_2 = 1$ and if $\sqrt{s}$ is of the form (7) with $v \in \mathcal{H}^{\beta}([0,1])$ and $g \in \mathcal{H}^{\alpha}\left([0, \|v\|_{\infty}^2]\right)$, then $\sqrt{s}$ is Hölderian with regularity $(\alpha, \alpha\beta)$ on $[0,1]^2$, and this regularity cannot be improved in general except in some particular situations. Under some smoothness assumption, the rate of estimation we would get is $n^{-2\alpha\beta/(2\alpha\beta+\beta+1)}$. This rate being always slower than the rate we obtain under the structural assumption on $\sqrt{s}$.

4. Families $\mathscr{F}$ of product functions.

A common way to model that the covariates influence the number of failures of n systems is to assume that, for each $i \in \{1, \ldots, n\}$, the intensity of N_i, is of the form $s(t, x_i) = u(t)v(x_i)$ where u is an unknown density function on $\mathbb{T}$, and v some unknown function from $\mathbb{X}$ into $\mathbb{R}_+$. This means, that in average, the number of failures of system i, $\mathbb{E}[N_i(\mathbb{T})] = v(x_i)$, depends on x_i through v only, and conditionally to $N_i(\mathbb{T}) = k_i > 0$, the times of failure are distributed along $\mathbb{T}$ independently of x_i, but accordingly to the density u.

We shall therefore consider the class $\mathscr{F}$ defined by

$$\mathscr{F} = \left\{\kappa v_1 v_2,\ \kappa \geq 0,\ (v_1, v_2) \in \mathbb{L}^2(\mathbb{T}, \mu) \times \mathbb{L}^2(\mathbb{X}, \nu_n),\ \|v_1\|_{\mathtt{t}} = \|v_2\|_{\mathtt{x}} = 1\right\}, \tag{9}$$

which amounts to assuming that s is of the form (or close to) a product function $u(t)v(x)$ with $u = v_1^2$ and $v = \kappa^2 v_2^2$.

In this section, we introduce collections of models $\mathbb{V}_1$ and $\mathbb{V}_2$ in order to approximate the components v_1 and v_2 separately. For each $V_1 \in \mathbb{V}_1$ to approximate v_1 and $V_2 \in \mathbb{V}_2$ to approximate v_2, we approximate $v_1 v_2$ by the model $V_1 \otimes V_2$ defined by

$$V_1 \otimes V_2 = \{v_1 v_2,\ (v_1, v_2) \in V_1 \times V_2\}. \tag{10}$$

The metric dimension of $V_1 \otimes V_2$ is controlled as follows.

Lemma 4.1. *Let V_1 and V_2 be a finite dimensional linear space of $\mathbb{L}^2(\mathbb{T}, \mu)$ and $\mathbb{L}^2(\mathbb{X}, \nu_n)$ respectively. The set $V_1 \otimes V_2$ defined by (10) has a finite metric dimension bounded by*

$$D_{V_1 \otimes V_2} = 1.4\,(\dim V_1 + \dim V_2 + 1).$$

By using Theorem 2.1, we prove the following result.

Proposition 4.1. *Let $\mathbb{V}_1$ and $\mathbb{V}_2$ be an at most countable collection of finite dimensional linear spaces of $\mathbb{L}^2(\mathbb{T},\mu)$ and $\mathbb{L}^2(\mathbb{X},\nu_n)$ respectively. Moreover, for $i\in\{1,2\}$, let Δ_i be some mapping on $\mathbb{V}_i$ with values into $[1,+\infty)$ such that $\sum_{V_i\in\mathbb{V}_i} e^{-\Delta_i(V_i)}\leq 1$.*

There exists an estimator $\hat{s}$ such that, for all $\kappa v_1 v_2\in\mathscr{F}$, where $\mathscr{F}$ is defined by (9), the following inequality holds:

$$\begin{aligned} C\mathbb{E}\left[H^2(s,\hat{s})\right] &\leq d_2^2(\sqrt{s},\kappa v_1 v_2)+\inf_{V_1\in\mathbb{V}_1}\left\{\kappa^2 d_{\mathtt{t}}^2(v_1,V_1)+\frac{\dim V_1+\Delta_1(V_1)}{n}\right\}\\ &\quad+\inf_{V_2\in\mathbb{V}_2}\left\{\kappa^2 d_{\mathtt{x}}^2(v_2,V_2)+\frac{\dim V_2+\Delta_2(V_2)}{n}\right\}\end{aligned}$$

where C is an universal positive contant. Furthermore, $\sqrt{\hat{s}}$ belongs to $\mathscr{F}$.

Apart for the term $d_2^2(\sqrt{s},\kappa v_1 v_2)$ which corresponds to some robustness with respect to the assumption $\sqrt{s}\in\mathscr{F}$, the risk bound we get corresponds to the one we would get if we could apply a model selection theorem on the components v_1 and v_2 separately.

4.1. **Smoothness assumptions on v_1 and v_2.** Let us illustrate this proposition by setting $\mathbb{T}=[0,1]^{k_1}$, $\mathbb{X}=[0,1]^{k_2}$, μ the Lebesgue measure and

$$\mathscr{F} = \left\{\kappa v_1 v_2, \kappa\geq 0, v_1\in\mathcal{H}([0,1]^{k_1}), \|v_1\|_{\mathtt{t}}=1, v_2\in\mathcal{H}([0,1]^{k_2}), \|v_2\|_{\mathtt{x}}=1\right\}. \tag{11}$$

Corollary 4.1. *There exists an estimator $\hat{s}$ such that, for all $\kappa v_1 v_2\in\mathscr{F}$, where $\mathscr{F}$ is defined by (11), the following inequality holds:*

$$\begin{aligned} C\mathbb{E}\left[H^2(s,\hat{s})\right] &\leq d_2^2(\sqrt{s},\kappa v_1 v_2)+\kappa^{\frac{2k_1}{2\bar{\alpha}+k_1}} L(v_1)^{\frac{2k_1}{2\bar{\alpha}+k_1}} n^{-\frac{2\bar{\alpha}}{2\bar{\alpha}+k_1}}\\ &\quad+\kappa^{\frac{2k_2}{2\bar{\beta}+k_2}} L(v_2)^{\frac{2k_2}{2\bar{\beta}+k_2}} n^{-\frac{2\bar{\beta}}{2\bar{\beta}+k_2}}+n^{-1}\end{aligned}$$

where $\boldsymbol{\alpha}\in(\mathbb{R}_+^\star)^{k_1}$, is such that $v_1\in\mathcal{H}^{\boldsymbol{\alpha}}([0,1]^{k_1})$, where $\boldsymbol{\beta}\in(\mathbb{R}_+^\star)^{k_2}$ is such that $v_2\in\mathcal{H}^{\boldsymbol{\beta}}([0,1]^{k_2})$, and where $C>0$ depends only on k_1, k_2, $\max_{1\leq i\leq k_1}\alpha_i$, and $\max_{1\leq i\leq k_2}\beta_i$.

In particular, if s is a product function of the form $\sqrt{s}=\kappa v_1 v_2$ for $v_1\in\mathcal{H}^{\boldsymbol{\alpha}}([0,1]^{k_1})$, and $v_2\in\mathcal{H}^{\boldsymbol{\beta}}([0,1]^{k_2})$, $\sqrt{s}$ is Hölderian with regularity $(\boldsymbol{\alpha},\boldsymbol{\beta})$ on $[0,1]^{k_1+k_2}$. However, the rate given by the corollary above is always faster than the one we would get by Corollary 3.1 under smoothness assumption only.

4.2. **Mixing smoothness and structural assumptions.** In Section 3.2.1 we have studied the case where $s(t,x)$ is of the form $\sqrt{s(t,x)}=g(t,<\theta_1,x>,\ldots,<\theta_l,x>)$ where g and $\theta_1,\ldots,\theta_l$ are unknown. We now assume that the covariate effect is multiplicative, and we approximate $\sqrt{s}$ by the class of functions of $\mathscr{F}$ defined by

$$\begin{aligned}\mathscr{F} &= \Big\{\kappa v_1 v_2,\ \kappa\geq 0,\ v_1\in\mathcal{H}([0,1]^{k_1}),\ \theta_1,\ldots,\theta_l\in\mathcal{B}(0,1), g\in\mathcal{H}([-1,1]^l),\\ &\qquad \forall x\in\mathbb{X},\ v_2(x)=g\left(<\theta_1,x>,\ldots,<\theta_l,x>\right),\ \|v_1\|_{\mathtt{t}}=\|v_2\|_{\mathtt{x}}=1\Big\}\end{aligned} \tag{12}$$

where we have chosen $\mathbb{T} = [0,1]^{k_1}$, μ the Lebesgue measure and $\mathbb{X} = \mathcal{B}(0,1) = \{x \in \mathbb{R}^{k_2}, \sum_{i=1}^{k_2} x_i^2 \leq 1\}$ the unit ball of $\mathbb{R}^{k_2}$.

Corollary 4.2. *There exists an estimator $\hat{s}$ such that, for all $\kappa v_1 v_2 \in \mathscr{F}$, where $\mathscr{F}$ is defined by (12) the following inequality holds:*

$$\begin{aligned} C\mathbb{E}\left[H^2(s,\hat{s})\right] \quad &\leq \quad d_2^2(\sqrt{s}, \kappa v_1 v_2) + \kappa^{\frac{2k_1}{2\bar{\alpha}+k_1}} L(v_1)^{\frac{2k_1}{2\bar{\alpha}+k_1}} n^{-\frac{2\bar{\alpha}}{2\bar{\alpha}+k_1}} \\ &\quad + \kappa^{\frac{2l}{2\bar{\beta}+l}} L(g)^{\frac{2l}{2\bar{\beta}+l}} n^{-\frac{2\bar{\beta}}{2\bar{\beta}+l}} + \frac{\ln\left(\kappa^2 \|g\|_{\boldsymbol{\beta}}^2 k_2^{-1}\right) \vee \ln n \vee 1}{n} k_2 \end{aligned}$$

where $\boldsymbol{\alpha} \in (\mathbb{R}_+^\star)^{k_1}$, $\boldsymbol{\beta} \in (\mathbb{R}_+^\star)^l$ are such that $v_1 \in \mathcal{H}^{\boldsymbol{\alpha}}([0,1]^{k_1})$, $g \in \mathcal{H}^{\boldsymbol{\beta}}([-1,1]^l)$ with $v_2(x) = g(<\theta_1, x>, \ldots, <\theta_l, x>)$ and where $C > 0$ depends only on k_1, l, $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$.

When $\sqrt{s}$ belongs to the class $\mathscr{F}$, the risk bound of the inequality above corresponds to the one we would get if we could estimate the functions v_1 and g separately. This risk bound is then better than the one we would get under smoothness assumptions only, or under the structural assumption of Section 3.2.1.

4.3. **A model selection theorem for parametric assumptions.** In this section, we introduce a parametric class $\mathscr{F}$ of the form

$$\mathscr{F} = \left\{a u_b v_\theta,\ a \geq 0,\ b \in I,\ \theta \in \mathbb{R}^{k_2}\right\},$$

where I is an interval of $\mathbb{R}$, $(u_b)_{b \in I}$ is a family of functions and v_θ is defined by $v_\theta(x) = \exp(<x, \theta>)$ for $x \in \mathbb{X} = \{x \in \mathbb{R}^{k_2}, \sum_{i=1}^{k_2} x_i^2 \leq 1\}$, the unit ball of $\mathbb{R}^{k_2}$. For each $i \in \{1, \ldots n\}$, the intensity of N_i, $s(\cdot, x_i)$, is thus assumed to be proportional to an element of (or an element close to) some reference parametric model $\{u_b^2,\ b \in I\}$. Let us give 3 examples of such models.

The Duane model (also known as the Power Law Processes) amounts to assuming that $u_b(t) = t^b$ with $b \in (-1/2, +\infty)$ and $t \in \mathbb{T} = (0,1]$. Proposed first in Duane (1964), this is one of the most popular models used in reliability. Indeed, although the intensity is simple, different situations can be modelled by this model. For example, if $b = 0$ each N_i obeys to an homogeneous Poisson process, whereas if $b > 0$ (respectively $b < 0$) the reliability of each system improves (respectively reduces) with time. In software reliability, we can cite the Goel-Okumoto model of Goel and Okumoto (1979) and the S-Shaped model of Yamada et al. (1983). The former amounts to assuming that $u_b(t)$ is of the form $u_b(t) = e^{-bt}$ whereas the latter corresponds to $u_b(t) = \sqrt{t}e^{-bt}$ where $b \in [0, +\infty)$ and $t \in \mathbb{T} = [0, +\infty)$.

A standard way to estimate s is to maximize the likelihood, as in Lawless (1987). However, our method is more general since we can (by Proposition 2.1) deal simultaneously with several (parametric and non parametric) assumptions. Moreover, the maximum likelihood estimator is not robust, which can be seen by adapting Section 2.3 of Birgé (2006) to Poisson processes.

To estimate s with our method, we have to consider the following assumption on the model $\{u_b,\, b \in I\}$.

Assumption 4.1 ($(u_b)_{b\in I}$). *The family $(u_b)_{b\in I}$ is a family of non vanishing functions of $\mathbb{L}^2(\mathbb{T},\mu)$ indexed by an interval I of the form $(b_0,+\infty)$. Moreover, there exists $\underline{\rho},\bar{\rho}>0$ two non increasing functions on I, such that for all $b,b'\in I$,*

$$\underline{\rho}\left(b\vee b'\right)|b-b'| \le \left\|\frac{u_b}{\|u_b\|_{\mathtt{t}}}-\frac{u_{b'}}{\|u_{b'}\|_{\mathtt{t}}}\right\|_{\mathtt{t}} \le \bar{\rho}\left(b\wedge b'\right)|b-b'|.$$

The purpose of the lemma below is to show that the assumption above holds for the Duane, Goel-Okumoto and S-Shaped models.

Lemma 4.2. *Let $I=(-1/2,+\infty)$, $\mathbb{T}=(0,1]$, μ the Lebesgue measure, and for $b\in I$, $u_b(t)=t^b$. Assumption 4.1 is satisfied with*

$$\forall u>-1/2, \quad \underline{\rho}(u)=\bar{\rho}(u)=\frac{1}{1+2u}.$$

Let $I=(0,+\infty)$, $\mathbb{T}=[0,+\infty)$, μ the Lebesgue measure, $k\in\mathbb{N}$, and for $b\in I$, $u_b(t)=t^{k/2}e^{-bt}$. Assumption 4.1 is satisfied with

$$\forall u>0, \quad \underline{\rho}(u)=\frac{1}{2u} \quad \text{and} \quad \bar{\rho}(u)=\frac{\sqrt{k+1}}{2u}.$$

To write our results, all along this section, $\|\cdot\|$ denotes the standard Euclidean norm of $\mathbb{R}^{k_2}$

$$\forall x\in\mathbb{R}^{k_2}, \quad \|x\|^2=\sum_{i=1}^{k_2} x_i^2$$

and d the distance induced by this norm.

Proposition 4.2. *Let $(u_b)_{b\in I}$ be a family such that Assumption 4.1 holds. There exist $\hat{a}\ge 0$, $\hat{b}\in I$ and $\hat{\theta}\in\mathbb{R}^{k_2}$, such that the estimator $\hat{s}=(\hat{a}u_{\hat{b}}v_{\hat{\theta}})^2$ satisfies, for all $a\ge 0$, $b\in I$, $\theta\in\mathbb{R}^{k_2}$, and $f\in\mathscr{F}$ of the form $f(t,x)=au_b(t)v_\theta(x)$,*

$$C\mathbb{E}\left[H^2(s,\hat{s})\right] \le d_2^2\left(\sqrt{s},f\right)+\frac{k_2\left(1\vee\|\theta\|\right)}{n}+\frac{C'}{n} \tag{13}$$

where C is an universal positive constant and where C' depends only on $\underline{\rho}$, $\bar{\rho}$, b_0 and b. More precisely,

$$C'=\ln\left[1\vee\bar{\rho}\left(b_0+\frac{b-b_0}{b-b_0+1}\right)\right]+\left|\ln\left(1\wedge\underline{\rho}(1+b)\right)\right|+\left|\ln(b-b_0)\right|.$$

Under parametric assumptions on s, this result says that the rate of convergence of $\hat{s}$ is of order n^{-1}, which is quite satisfying when n is large, but may be inadequate in a non-asymptotic point of view. Indeed, the second term of the right-hand side of inequality (13) may be large especially when k_2 is large, says larger than n. To avoid this difficulty, a convenient assumption amounts to considering that θ is sparse, which means that θ is close to some (unknown) linear subspace W of $\mathbb{R}^{k_2}$ with $\dim W$ small. We generalize afterwards Proposition 4.2 to take account of this situation.

Proposition 4.3. *Let $(u_b)_{b\in I}$ be a family such that Assumption 4.1 holds. Let $\mathbb{W}$ be an at most countable family of linear subspaces of $\mathbb{R}^{k_2}$ and let Δ be a non-negative map on $\mathbb{W}$ such that $\sum_{W\in\mathbb{W}} e^{-\Delta(W)} \leq 1$.*

There exist $\hat{a} \geq 0$, $\hat{b} \in I$ and $\hat{\theta} \in \mathbb{R}^{k_2}$, such that the estimator $\hat{s} = (\hat{a} u_{\hat{b}} v_{\hat{\theta}})^2$ satisfies, for all $a \geq 0$, $b \in I$, $\theta \in \mathbb{R}^{k_2}$, and $f \in \mathscr{F}$ of the form $f(t,x) = a u_b(t) v_\theta(x)$,

$$\begin{aligned} C\mathbb{E}\left[H^2(s,\hat{s})\right] &\leq d_2^2(\sqrt{s}, f) + \frac{C'}{n} \\ &\quad + \inf_{W\in\mathbb{W}} \left\{ a^2 \|u_b\|_{\mathfrak{t}}^2 e^{2\|\theta\|} d^2(\theta, W) + \frac{(1\vee \dim W)(1 \vee \|\theta\|) + \Delta(W)}{n} \right\} \end{aligned}$$

where C is an universal positive constant and where C' is given by Proposition 4.2.

For illustration purpose, let us make explicit the constant C' for the Duane model, and let us therefore assume that there exist some unknown parameters a, b, θ such that s is of the form $\sqrt{s(t,x)} = a t^b \exp(< \theta, x >)$. We derive from Proposition 4.2 an estimator whose risk satisfies

$$C\mathbb{E}\left[H^2(s,\hat{s})\right] \leq \frac{(1\vee \|\theta\|)\, k_2 + |\ln(2b+1)|}{n} \tag{14}$$

where C is an universal positive constant. However, if for instance k_2 is large and if most of the components of θ are small or null, the preceding proposition can be used to improve substantially the risk of our estimators. For simplicity, assume that

$$k_\star = |\{i \in \{1,\dots,k_2\}, \theta_i \neq 0\}|$$

is small. We then define $\mathcal{M}$ the set of all subsets of $\{1,\dots,k_2\}$, and for each $m \in \mathcal{M}$, the set

$$W_m = \{(y_1,\dots,y_{k_2}),\, \forall i \notin m,\, y_i = 0\} \subset \mathbb{R}^{k_2}.$$

We apply Proposition 4.3 with

$$\mathbb{W} = \{W_m,\, m \in \mathcal{M}\} \quad \text{and} \quad \forall m \in \mathcal{M},\ \Delta(W_m) = 1 + |m| + \ln\binom{k_2}{|m|}.$$

This leads to an estimator $\hat{s}$ such that

$$C\mathbb{E}\left[H^2(s,\hat{s})\right] \leq \frac{(1 \vee \ln k_2 \vee \|\theta\|)(1\vee k_\star) + |\ln(2b+1)|}{n},$$

which improves inequality (14) when $k_\star$ is small and k_2 large.

5. Parametric models.

In this section, we consider the situation where the intensity of each process N_i belongs (or is close) to a same parametric model. Let us define Θ a subset of $\mathbb{R}^k$, and let us denote by $\mathcal{F} = \{f_{\boldsymbol{\theta}},\, \boldsymbol{\theta} \in \Theta\}$ a class of functions of $\mathbb{L}^2(\mathbb{T},\mu)$. Our aim is to estimate s when, for each $i \in \{1,\dots n\}$, the square root of the intensity of the Poisson process N_i, $\sqrt{s(\cdot, x_i)}$, is (or is close to) an element of $\mathcal{F}$. We introduce thus the class of functions $\mathscr{F}$ defined by

$$\mathscr{F} = \left\{(t,x) \mapsto f_{u(x)}(t), \text{ where } u \text{ is a map on } \mathbb{X} \text{ with values into } \Theta\right\}.$$

For instance, if $\mathcal{F}$ corresponds to the Duane model (see Section 4.3), we assume that there exist two functions a and $b > -1/2$ on $\mathbb{X}$ such that $\sqrt{s(t,x)}$ is close to a function of the form $a(x)t^{b(x)}$. The class $\mathscr{F}$ is then the set of all functions of this form. For more general classes $\mathcal{F}$, we have to consider the following assumption.

Assumption 5.1. *There exists $\mathcal{K}$ an at most countable collection of closed convex subsets of $\mathbb{R}^k$ such that $\Theta = \cup_{K \in \mathcal{K}} K$. Moreover, for each $K \in \mathcal{K}$, there exist $\boldsymbol{\alpha}(K) = (\alpha_j(K))_{1 \le j \le k} \in (0,1]^k$ and $\boldsymbol{R}(K) = (R_j(K))_{1 \le j \le k} \in (\mathbb{R}_+^\star)^k$ such that*

$$\forall \boldsymbol{\theta}, \boldsymbol{\theta}' \in K, \quad \|f_{\boldsymbol{\theta}} - f_{\boldsymbol{\theta}'}\|_{\mathtt{t}} \le \sum_{j=1}^{k} R_j(K)\,|\theta_j - \theta'_j|^{\alpha_j(K)}. \tag{15}$$

Typically, relation (15) does not hold for $K = \Theta$ but rather for some subsets K of Θ. In Sections 5.1, 5.2 and 5.3, we give some examples of processes satisfying the assumption above. We assume in these sections that μ is the Lebesgue measure.

5.1. **The Duane model.** As explained previously, this model is defined by $\mathcal{F} = \{f_{\boldsymbol{\theta}},\, \boldsymbol{\theta} \in \Theta\}$ where $\Theta = \mathbb{R} \times (-1/2, +\infty)$, and where $f_{\boldsymbol{\theta}}$ is of the form $f_{\boldsymbol{\theta}}(t) = \theta_1 t^{\theta_2}$ for $\boldsymbol{\theta} = (\theta_1, \theta_2) \in \Theta$ and $t \in \mathbb{T} = (0,1]$. The class $\mathscr{F}$ is thus

$$\mathscr{F} = \left\{(t,x) \mapsto a(x)t^{b(x)}, \text{where } a : \mathbb{X} \to \mathbb{R} \text{ and } b : \mathbb{X} \to (-1/2, +\infty)\right\}.$$

In this situation, the lemma below gives a collection $\mathcal{K}$ such that Assumption 5.1 holds.

Lemma 5.1. *For all $i_1, i_2 \in \mathbb{N}^\star$, let $K_{i_1,i_2} = [-i_1, i_1] \times [-\frac{1}{2} + \frac{1}{i_2}, +\infty)$. Assumption 5.1 is satisfied for the Duane model with $\mathcal{K} = \{K_{i_1,i_2},\, i_1, i_2 \in \mathbb{N}^\star\}$, and with*

$$\boldsymbol{\alpha}\left(K_{i_1,i_2}\right) = (1,1) \quad \textit{and} \quad \mathbf{R}\left(K_{i_1,i_2}\right) = \left(i_2^{1/2}, \sqrt{2}\, i_1 i_2^{3/2}\right).$$

5.2. **The Goel-Okumoto and S-Shaped models.** As in Section 4.3, we study these two models simultaneously. For $k \in \mathbb{N}^\star$ and $\boldsymbol{\theta} = (\theta_1, \theta_2) \in \Theta = \mathbb{R}_+ \times \mathbb{R}_+^\star$ we set $f_{\boldsymbol{\theta}}(t) = \sqrt{\theta_1 t^{k-1} e^{-\theta_2 t}}$ for $t \in \mathbb{T} = (0, +\infty)$ and

$$\mathcal{F} = \left\{t \mapsto \sqrt{\theta_1 t^{k-1} e^{-\theta_2 t}},\, (\theta_1, \theta_2) \in \mathbb{R}_+ \times \mathbb{R}_+^\star\right\}. \tag{16}$$

The class $\mathscr{F}$ is thus

$$\mathscr{F} = \left\{(t,x) \mapsto \sqrt{a(x) t^{k-1} e^{-b(x)t}}, \text{where } a : \mathbb{X} \to \mathbb{R}_+ \text{ and } b : \mathbb{X} \to \mathbb{R}_+^\star\right\}.$$

Lemma 5.2. *For all $i_1, i_2 \in \mathbb{N}^\star$, let $K_{i_1,i_2} = [0, i_1] \times [\frac{1}{i_2}, +\infty)$. Assumption 5.1 is satisfied for the class $\mathcal{F}$ defined by (16) with $\mathcal{K} = \{K_{i_1,i_2},\, i_1, i_2 \in \mathbb{N}^\star\}$, and with*

$$\boldsymbol{\alpha}\left(K_{i_1,i_2}\right) = (1/2, 1/2) \quad \textit{and} \quad \mathbf{R}\left(K_{i_1,i_2}\right) = \left(\sqrt{(k-1)!\, i_2^k}, \sqrt{k!\, i_1 i_2^{k+1}}\right).$$

5.3. **Translations and dilatations from a reference model.** Let φ be a function defined on an interval I of $\mathbb{R}$. We assume that $\mathbb{T}$ is an interval and we define J_2 and J_3 two intervals such that, for all $t \in \mathbb{T}$, $\theta_2 \in J_2$ and $\theta_3 \in J_3$, $t\theta_2 - \theta_3$ belongs to I. In this section, we study the situation in which the class $\mathscr{F}$ gathers the functions of the form

$$(t, x) \mapsto a_1(x)\varphi\left(ta_2(x) - a_3(x)\right)$$

where a_1, a_2, a_3 are real-valued functions on $\mathbb{X}$ such that $a_2(\mathbb{X}) \subset J_2$ and $a_3(\mathbb{X}) \subset J_3$. This class corresponds to several parametric models $\mathcal{F}$. For instance the Musa-Okumoto model (see Musa and Okumoto (1984)) corresponds to $\varphi(t) = 1/\sqrt{1+t}$ and the Cox Lewis processes (see Cox and Lewis (1966)) to $\varphi(t) = e^t$. The case where there exist three *constant* functions a_1, a_2, a_3 such that $\sqrt{s(t,x)} = a_1\varphi\left(ta_2(x) - a_3(x)\right)$ has been dealt in several papers. We refer for instance to Dabye (1993), Aubry and Dabye (2001) and the references therein for estimating s by using a maximum likelihood estimator or a minimum distance estimator.

We consider the following assumption on φ.

Assumption 5.2. *For all $\theta_2 \in J_2$ and $\theta_3 \in J_3$, the function $t \mapsto \varphi(t\theta_2 - \theta_3)$ belongs to $\mathbb{L}^2(\mathbb{T}, \mu)$. Moreover, there exists $\varrho_2, \varrho_3 > 0$ and $0 < \beta, \gamma \leq 1$, such that for all $\theta_2, \theta_2' \in J_2$ and all $\theta_3, \theta_3' \in J_3$,*

$$\int_{\mathbb{T}} \left(\varphi\left(t\theta_2 - \theta_3\right) - \varphi\left(t\theta_2' - \theta_3'\right)\right)^2 \mathrm{d}\mu(t) \leq \varrho_2^2 \left|\theta_2 - \theta_2'\right|^{2\beta} + \varrho_3^2 |\theta_3 - \theta_3'|^{2\gamma}.$$

This assumption is satisfied in particular if $\mathbb{T}$ is bounded, and if φ admits a bounded differential on $I \setminus I'$ where I' is at most countable.

Lemma 5.3. *Assume that Assumption 5.2 holds. Let $\Theta = \mathbb{R} \times J_2 \times J_3$ and for $\boldsymbol{\theta} = (\theta_1, \theta_2, \theta_3) \in \Theta$ and $t \in \mathbb{T}$, let $f_{\boldsymbol{\theta}}(t) = \theta_1\varphi(t\theta_2 - \theta_3)$. We then define for $i \in \mathbb{N}^\star$, $K_i = [-i, i] \times J_2 \times J_3$. Assumption 5.1 holds with $\mathcal{K} = \{K_i,\, i \in \mathbb{N}^\star\}$, $\boldsymbol{\alpha}(K_i) = (1, \beta, \gamma)$ and*

$$\mathbf{R}(K_i) = \left(\sqrt{2} \sup_{(\theta_2,\theta_3) \in J_2 \times J_3} \left(\int_{\mathbb{T}} \varphi^2(\theta_2 t - \theta_3)\, \mathrm{d}\mu(t)\right)^{1/2}, \sqrt{2} i \varrho_2, \sqrt{2} i \varrho_3\right).$$

5.4. **A model selection theorem.** The main theorem of this section is the following.

Theorem 5.1. *Assume that Assumption 5.1 holds. Let $\mathbb{W}_1, \ldots, \mathbb{W}_k$ be k families of finite dimensional linear spaces of $\mathbb{L}^2(\mathbb{X}, \nu_n)$. Let $\Delta_{\mathcal{K}}$ be some non-negative mapping on $\mathcal{K}$ such that $\sum_{K \in \mathcal{K}} e^{-\Delta_{\mathcal{K}}(K)} \leq 1$, and for each $1 \leq j \leq k$, Δ_j be a non-negative mapping on $\mathbb{W}_j$ such that $\sum_{W_j \in \mathbb{W}_j} e^{-\Delta_j(W_j)} \leq 1$.*

One can build an estimator $\hat{s}$ such that for all $K \in \mathcal{K}$, all map u from $\mathbb{X}$ with values into K, and $f \in \mathscr{F}$ of the form $f(t, x) = f_{u(x)}(t)$,

$$C\mathbb{E}\left[H^2(s, \hat{s})\right] \leq d_2^2(\sqrt{s}, f) + \sum_{j=1}^{k} \varepsilon_j(u_j) + \frac{\Delta_{\mathcal{K}}(K)}{n},$$

where $\varepsilon_j(u_j)$ is defined by

$$\varepsilon_j(u_j) = \inf_{W_j \in \mathbb{W}_j} \left\{ (R_j(K))^2 \; d_{\mathtt{x}}^{2\alpha_j(K)}(u_j, W_j) + \frac{(\dim(W_j) \vee 1)\, \tau_{u,j}(n) + \Delta_j(W_j)}{n} \right\},$$

where

$$\tau_{u,j}(n) = 1 + \ln n + \ln(1 \vee R_j(K)) + \ln(1 \vee \|u_j\|_{\mathtt{x}}),$$

and where $C > 0$ depends only on k and the maps $\alpha_1, \dots, \alpha_k$.

Roughly speaking, this result says that the risk bound we get when $\sqrt{s}$ is of the form $\sqrt{s(t,x)} = f_{u(x)}(t)$, corresponds to the one we would get if we could apply a model selection theorem on the components $u_1, \dots, u_k$ separately. As in Section 3, each term $\varepsilon_j(u_j)$ can be controlled under some structural or smoothness assumptions on u_j. For instance, if $\mathbb{X} = [0,1]^{k_2}$ and if u_j is assumed to belong to the class $\mathscr{F}_j = \mathcal{H}([0,1]^{k_2})$, by a similar argument as the one used in the proof of Corollary 3.1, $\varepsilon_j(u_j)$ is such that

$$C_j \varepsilon_j(u_j) \leq \left(R_j(K) L(u_j)^{\alpha_j(K)} \right)^{\frac{2k_2}{k_2 + 2\alpha_j(K)\bar{\boldsymbol{\beta}}_j}} \left(\frac{\tau_{u,j}(n)}{n} \right)^{\frac{2\alpha_j(K)\bar{\boldsymbol{\beta}}_j}{2\alpha_j(K)\bar{\boldsymbol{\beta}}_j + k_2}} + \frac{\tau_{u,j}(n)}{n}$$

where $\boldsymbol{\beta}_j$ is such that $u_j \in \mathcal{H}^{\boldsymbol{\beta}_j}([0,1]^{k_2})$ and where $C_j > 0$ depends only on k_2 and $\boldsymbol{\beta}_j$. In particular, if $\alpha_j(K) = 1$ and if n is large, $\varepsilon_j(u_j)$ is of order $(\ln n / n)^{2\bar{\boldsymbol{\beta}}_j/(2\bar{\boldsymbol{\beta}}_j + k_2)}$. Apart from the logarithmic factor, this corresponds to the estimation rate of an Hölderian function on $[0,1]^{k_2}$. The corollary below illustrates this result in the framework of the Duane model.

Corollary 5.1. *One can build an estimator $\hat{s}$ such that, for all $\boldsymbol{\alpha} \in \left(\mathbb{R}_+^\star\right)^{k_2}$, $\boldsymbol{\beta} \in (\mathbb{R}_+^\star)^{k_2}$, for all $a \in \mathcal{H}^{\boldsymbol{\alpha}}([0,1]^{k_2})$, $b \in \mathcal{H}^{\boldsymbol{\beta}}([0,1]^{k_2})$ satisfying $b > -1/2$, and for a function f of the form $f(t,x) = a(x) t^{b(x)}$,*

$$\begin{aligned} C\mathbb{E}\left[H^2(s,\hat{s})\right] \quad \leq \quad & d_2^2\left(\sqrt{s}, f\right) + \left(\frac{1}{1 \wedge \inf_{x \in [0,1]^{k_2}} (2b(x)+1)} \right)^{\frac{2k_2}{2\bar{\boldsymbol{\alpha}} + k_2}} L(a)^{\frac{2k_2}{2\bar{\boldsymbol{\alpha}} + k_2}} \left(\frac{\ln n}{n} \right)^{\frac{2\bar{\boldsymbol{\alpha}}}{2\bar{\boldsymbol{\alpha}} + k_2}} \\ & + \left(\frac{1 \vee \|a\|_\infty}{1 \wedge \inf_{x \in [0,1]^{k_2}} (2b(x)+1)^3} \right)^{\frac{2k_2}{2\bar{\boldsymbol{\beta}} + k_2}} L(b)^{\frac{2k_2}{2\bar{\boldsymbol{\beta}} + k_2}} \left(\frac{\ln n}{n} \right)^{\frac{2\bar{\boldsymbol{\beta}}}{2\bar{\boldsymbol{\beta}} + k_2}} \\ & + C' \frac{1 \vee \ln n}{n} \end{aligned}$$

where $C > 0$ depends on k_2, $\max_{1 \leq j \leq k} \alpha_j$, $\max_{1 \leq j \leq k} \beta_j$, and where C' depends on $L(a)$, $L(b)$, $\bar{\boldsymbol{\alpha}}$, $\bar{\boldsymbol{\beta}}$, $\|a\|_\infty$, $\|b\|_\infty$ and $\inf_{x \in [0,1]^{k_2}} (2b(x)+1)$.

5.5. **Change point detection.** In the case where the intensity s_i of each N_i is of the form $\sqrt{s_i(t)} = f_{\boldsymbol{\theta}_i}(t)$, a natural way to control the risk of our estimator $\hat{s}$ is to consider some assumptions on the map $i \mapsto \boldsymbol{\theta}_i$. This problem amounts to choosing suitable collections $\mathbb{W}_1, \dots, \mathbb{W}_k$ to approximate functions on $\mathbb{X} = \{1, \dots, n\}$.

In this section, we focus on the case where the map $i \mapsto \boldsymbol{\theta}_i$ is piecewise constant with a small number of jumps, and we assume that $n \geq 2$. We introduce the set $\mathcal{P}$ of partitions of $\{1, \dots, n\}$ into intervals and aim at estimating s when it is of the form

$$\forall (t,i) \in \mathbb{T} \times \{1, \dots, n\}, \quad \sqrt{s(t,i)} = \sum_{I \in P} f_{\boldsymbol{\theta}_I}(t) \mathbb{1}_I(i), \tag{17}$$

where $P \in \mathcal{P}$ is an unknown partition and $(\boldsymbol{\theta}_I)_{I \in P}$ an unknown family of elements of Θ.

We define for each partition $P \in \mathcal{P}$, the linear space of piecewise constants

$$W_P = \left\{ \sum_{I \in P} a_I \mathbb{1}_I, \, a_I \in \mathbb{R} \right\}$$

and apply Theorem 5.1 with the collections and maps defined by

$$\forall j \in \{1, \dots, k\}, \; \mathbb{W}_j = \{W_P, \, P \in \mathcal{P}\} \; \text{ and } \Delta_j(W_P) = 1 + |P| + \ln \binom{n-1}{|P|-1}.$$

This leads to the result below.

Corollary 5.2. *Assume that Assumption 5.1 and relation (17) hold. Let $\Delta_{\mathcal{K}}$ be a non-negative map on $\mathcal{K}$ such that $\sum_{K \in \mathcal{K}} e^{-\Delta(K)} \leq 1$.*

One can build an estimator $\hat{s}$ such that,

$$C\mathbb{E}\left[H^2(s, \hat{s})\right] \leq |P| \frac{\ln n + C'}{n} + \frac{\Delta_{\mathcal{K}}(K)}{n},$$

where K is such that $\boldsymbol{\theta}_I \in K$ for all $I \in P$, where $C > 0$ depends only on k and $\alpha_1, \dots, \alpha_k$, and where C' is given by

$$C' = 1 + \sup_{1 \leq j \leq k} \left(\ln\left(1 + R_j(K)\right)\right) + \sup_{I \in P} \left(\ln\left(1 + \|\boldsymbol{\theta}_I\|_\infty\right)\right),$$

where $\|\boldsymbol{\theta}_I\|_\infty = \sup_{1 \leq j \leq k} |(\boldsymbol{\theta}_I)_j|$.

For illustration purpose, in the context of the Duane model (see Section 5.1), there exist $a_1 \dots, a_n \in \mathbb{R}$, and $b_1, \dots, b_n \in (-1/2, +\infty)$ such that $\sqrt{s_i(t)} = a_i t^{b_i}$, whatever $t \in (0,1]$. The preceding corollary provides then an estimator $\hat{s}$ such that

$$C\mathbb{E}\left[H^2(s, \hat{s})\right] \leq (1 + r_1 + r_2) \frac{\ln n + C'}{n}$$

where r_1 and r_2 are the numbers of jumps of the maps $i \mapsto a_i$ and $i \mapsto b_i$ respectively, where C is an universal positive constant, and where C' depends on $\sup_{1 \leq i \leq n} |a_i|$, $\sup_{1 \leq i \leq n} |b_i|$ and $\inf_{1 \leq i \leq n} (2b_i + 1)$.

The preceding collections $\mathbb{W}_1, \dots, \mathbb{W}_k$ can also be used to approximate the map $i \mapsto \boldsymbol{\theta}_i$ under others assumptions such as smoothness ones. For instance, an approximation theorem for monotone functions can be found in Baraud and Birgé (2009) and can be used to deal with the situation where some components of the map $i \mapsto \boldsymbol{\theta}_i$ are monotone.

6. Proofs

6.1. **Proofs of Section 2.**

6.1.1. *Proof of Theorem 2.1.* Let us introduce the following measure N on the measurable space $(\mathbb{T}\times\mathbb{X},\mathcal{T}\otimes\mathcal{X})$

$$\forall A\in\mathcal{T}\otimes\mathcal{X},\quad N(A)=\frac{1}{n}\sum_{i=1}^{n}\left(N_i\otimes\delta_{x_i}\right)(A).$$

This measure is related to M (defined in Section 1) by the relation

$$\forall A\in\mathcal{T}\otimes\mathcal{X},\ \mathbb{E}\left[N(A)\right]=\int_A s(u)\,\mathrm{d}M(u).$$

The problem of estimating such a s from two observed measures (N,M) has been considered in Baraud and Birgé (2009); Baraud (2010). As explained in the introduction, our approach is based on robust testing, and we borrow the test of Baraud (2010). This test is a function T on $\mathbb{L}^1_+(\mathbb{T}\times\mathbb{X},M)\times\mathbb{L}^1_+(\mathbb{T}\times\mathbb{X},M)$ with values into $\mathbb{R}$. For two functions s_1 and s_2 of $\mathbb{L}^1_+(\mathbb{T}\times\mathbb{X},M)$, it is defined by

$$\begin{aligned}
T(s_1,s_2) &= \frac{1}{2n}\sum_{i=1}^{n}\int_{\mathbb{T}}\sqrt{\frac{s_1(t,x_i)+s_2(t,x_i)}{2}}\left(\sqrt{s_2(t,x_i)}-\sqrt{s_1(t,x_i)}\right)\mathrm{d}\mu(t)\\
&\quad+\frac{1}{\sqrt{2}n}\sum_{i=1}^{n}\int_{\mathbb{T}}\frac{\sqrt{s_2(t,x_i)}-\sqrt{s_1(t,x_i)}}{\sqrt{s_1(t,x_i)+s_2(t,x_i)}}\,\mathrm{d}N_i(t)\\
&\quad-\frac{1}{2n}\sum_{i=1}^{n}\int_{\mathbb{T}}\left(s_2(t,x_i)-s_1(t,x_i)\right)\,\mathrm{d}\mu(t).
\end{aligned}$$

Given two functions s_1 and s_2 this test returns the one deemed as being the one the closest to s. If $T(s_1,s_2)>0$, we prefer s_2 to s_1 and if $T(s_1,s_2)<0$ we prefer s_1 to s_2. If $T(s_1,s_2)=0$ we prefer arbitrary s_1 or s_2.

We shall derive a T-estimator $\hat{s}$ from this test by using the device of Birgé (2006). In order to do so, we have to verify that this test is robust. This is the purpose of the lemma below, whose proof is postponed to Section 6.1.2.

Lemma 6.1. *There exist $\kappa_0>0$ and a positive function a on $(\kappa_0,+\infty)$, such that for all $\kappa>\kappa_0$, for all $x\in\mathbb{R}$, and all $s_1,s_2\in\mathbb{L}^1_+(\mathbb{T}\times\mathbb{X},M)$ satisfying $\kappa H(s,s_1)\leq H(s_1,s_2)$,*

$$\mathbb{P}\left[T(s_1,s_2)\geq x\right]\leq\exp\left[-na(\kappa)\left(H^2(s_1,s_2)+x\right)\right].$$

We now derive from the collection $\mathbb{V}$ a family of D-models (in sense of Definition 4 of Birgé (2006)) for the metric space $(\mathbb{L}^1_+(\mathbb{T}\times\mathbb{X}),M)$ by the lemma below.

Lemma 6.2. *For all $\eta>0$, there exists an at most countable collection $\{T_V(\eta),\ V\in\mathbb{V}\}$, such that each $T_V(\eta)$ is a D-model of the metric space $(\mathbb{L}^1_+(\mathbb{T}\times\mathbb{X},M),H)$ with parameters η, $63D_V(\eta/2)$ and 1. Moreover,*

$$H(s,T_V(\eta))\leq 2\sqrt{2}\left(d_2\left(\sqrt{s},V\right)+\eta\right), \tag{18}$$

and for all $f \in T_V(\eta)$*, there exists* $g \in V$ *such that* $\sqrt{f} = g \vee 0$*.*

Proof. Thanks to Proposition 7 of Birgé (2006), we can define for each $V \in \mathbb{V}$ and $\eta > 0$ a D-model $S'_V(\eta) \subset V$ of $\mathbb{L}^2(\mathbb{T} \times \mathbb{X}, M)$ with parameters η, $7D_V(\eta/2)$ and 1, such that

$$\forall f \in V, \quad d_2(f, S'_V(\eta)) \leq \eta.$$

We apply for each $V \in \mathbb{V}$, Proposition 12 of Birgé (2006), with $T = S'_V(\eta)$, $M_0 = \mathbb{L}^2_+(\mathbb{T} \times \mathbb{X}, M)$, $M' = \mathbb{L}^2(\mathbb{T} \times \mathbb{X}, M)$ and $\bar{\pi}$ defined by $\bar{\pi}(f) = f \vee 0$. This gives a subset $S''_V(\eta) \subset \mathbb{L}^2_+(\mathbb{T} \times \mathbb{X}, M)$ such that

$$\forall f \in \mathbb{L}^2(\mathbb{T} \times \mathbb{X}, M), \forall x \geq 2, \quad |S''_V(\eta) \cap \mathcal{B}(f, x\eta)| \leq \exp\left(63 D_V(\eta)\, x^2\right)$$

where $\mathcal{B}(f, x\eta)$ is the ball centered at f with radius $x\eta$ of the metric space $(\mathbb{L}^2(\mathbb{T} \times \mathbb{X}, M), d_2)$, and such that

$$\forall f \in \mathbb{L}^2_+(\mathbb{T} \times \mathbb{X}, M), \quad d_2(f, S''_V(\eta)) \leq 4 d_2(f, S'_V(\eta)).$$

The lemma holds with $T_V(\eta) = \{f^2,\, f \in S''_V(\eta)\}$. □

With no loss of generality, we can assume that D_V is non-increasing and that $a(2\kappa_0) \leq 1$ where a and κ_0 are given in Lemma 6.1. Let us set

$$\eta'_V = \frac{21}{2}\sqrt{\frac{3}{5a(2\kappa_0)}}\eta_V.$$

For all $\eta \geq 2\eta'_V$, $T_V(\eta)$ is a D-model with parameters η, $63D_V(\eta_V)$ and 1. It follows from Theorem 5 of Birgé (2006) (applied with $\mathcal{M} = \mathbb{V}$, $S = \cup_{V \in \mathbb{V}} T_V(2\eta'_V \vee \sqrt{21 n^{-1}(a(2\kappa_0))^{-1}\Delta(V)})$, $D_m = 63 D_V(\eta_V)$, $\eta_m^2 = (2\eta'_V)^2 \vee 21 n^{-1}(a(2\kappa_0))^{-1}\Delta(V)$), that there exists

$$\hat{s} \in \bigcup_{V \in \mathbb{V}} T_V\left((2\eta'_V)^2 \vee 21 n^{-1}(a(2\kappa_0))^{-1}\Delta(V)\right)$$

such that

$$C\mathbb{E}\left[H^2(s, \hat{s})\right] \leq \inf_{V \in \mathbb{V}} \left\{ H^2\left(s, T_V\left((2\eta'_V)^2 \vee 21 n^{-1}(a(2\kappa_0))^{-1}\Delta(V)\right)\right) + \eta'^2_V + \frac{\Delta(V)}{n} \right\},$$

where C is an universal positive constant. By using inequality (18),

$$C'\mathbb{E}\left[H^2(s, \hat{s})\right] \leq \inf_{V \in \mathbb{V}} \left\{ d_2^2\left(\sqrt{s}, V\right) + \eta'^2_V + \frac{\Delta(V)}{n} \right\}.$$

The conclusion follows. □

6.1.2. *Proof of Lemma 6.1.* We start with the following concentration inequality.

Lemma 6.3. *Let $f_1, \dots, f_n$ be n measurable bounded functions satisfying*

$$\frac{1}{n}\sum_{i=1}^{n}\int_{\mathbb{T}} f_i^2(t)s_i(t)\,\mathrm{d}\mu(t) \leq \upsilon.$$

The following inequality holds for all $r \geq 0$:

$$\begin{aligned}\mathbb{P}\left(\frac{1}{n}\sum_{i=1}^{n}\int_{\mathbb{T}} f_i(t)(\,\mathrm{d}N_i(t) - s_i(t)\,\mathrm{d}\mu(t)) \geq r\right) &\leq \exp\left(-n\frac{\upsilon}{b^2}h\left(\frac{\rho r}{\upsilon}\right)\right)\\ &\leq \exp\left(-n\frac{r^2}{2\left(\upsilon + \frac{\rho r}{3}\right)}\right)\end{aligned}$$

where $\rho = \max_{1\leq i\leq n}\|f_i\|_\infty$ and where h is the function defined for $u \in (-1, +\infty)$ by $h(u) = (1+u)\ln(1+u) - u$.

Proof. By homogeneity we can assume that $\rho = 1$. We assume moreover that for each $i \in \{1, \dots, n\}$, f_i is a piecewise constant function. There exist thus $k_1, \dots, k_n \in \mathbb{N}^\star$ and a family $(a_{i,j})_{\substack{1\leq i\leq n\\ 1\leq j\leq k_i}}$ of elements of $[-1, 1]$ such that

$$\forall t \in \mathbb{T}, \quad f_i(t) = \sum_{j=1}^{k_i} a_{i,j}\mathbb{1}_{A_{i,j}}(t)$$

where the $A_{i,j}$ are measurable sets of $\mathbb{T}$ such that $A_{i,j} \cap A_{i,j'} = \emptyset$ for all $j \neq j'$. We have for all $t \geq 0$:

$$\begin{aligned}\ln\mathbb{E}\left(e^{t\sum_{i=1}^{n}\left[\int_{\mathbb{T}} f_i\,\mathrm{d}N_i - \mathbb{E}\left(\int_{\mathbb{T}} f_i\,\mathrm{d}N_i\right)\right]}\right) &= \sum_{i=1}^{n}\ln\mathbb{E}\left(e^{t\left[\int_{\mathbb{T}} f_i\,\mathrm{d}N_i - \mathbb{E}\left(\int_{\mathbb{T}} f_i\,\mathrm{d}N_i\right)\right]}\right)\\ &= \sum_{i=1}^{n}\sum_{j=1}^{k_i}\ln\mathbb{E}\left(e^{ta_{i,j}[N_i(A_{i,j}) - \mathbb{E}(N_i(A_{i,j}))]}\right)\\ &= \sum_{i=1}^{n}\sum_{j=1}^{k_i}\mathbb{E}\left(N_i(A_{i,j})\right)\left(e^{ta_{i,j}} - ta_{i,j} - 1\right).\end{aligned}$$

By using the monotony of the function $x \mapsto (e^x - x - 1)/x^2$,

$$\begin{aligned}\ln\mathbb{E}\left(e^{t\sum_{i=1}^{n}\left[\int_{\mathbb{T}} f_i\,\mathrm{d}N_i - \mathbb{E}\left(\int_{\mathbb{T}} f_i\,\mathrm{d}N_i\right)\right]}\right) &\leq \sum_{i=1}^{n}\sum_{j=1}^{k_i}\mathbb{E}\left(a_{i,j}^2 N_i(A_{i,j})\right)\left(e^t - t - 1\right)\\ &\leq \upsilon\left(e^t - t - 1\right).\end{aligned}$$

This result still holds when f is not piecewise constant by using the fact that a measurable function can be approximated by piecewise constant functions. The concentration inequality is then deduced from Cramér-Chernoff method, see Chapter 2 of Massart (2003).

□

Let us define the function ψ on $\mathbb{R}_+^2$ with values into $\left[-1/\sqrt{2}, 1/\sqrt{2}\right]$ by

$$\forall t_1, t_2 \in \mathbb{R}_+, \quad \psi(t_1, t_2) = \frac{1}{\sqrt{2}}\left(\sqrt{\frac{1}{1+t_1/t_2}} - \sqrt{\frac{1}{1+t_2/t_1}}\right),$$

where we use the convention $0/0 = 0$ and $a/\infty = 0$ for all $a \geq 0$. Let then

$$\begin{aligned} Z(s_1, s_2) &= T(s_1, s_2) - \mathbb{E}\left[T(s_1, s_2)\right] \\ &= \int_{\mathbb{T}\times\mathbb{X}} \psi(s_1, s_2)\, \mathrm{d}M - \mathbb{E}\left(\int_{\mathbb{T}\times\mathbb{X}} \psi(s_1, s_2)\, \mathrm{d}M\right). \end{aligned}$$

We derive from Corollary 1 of Baraud (2010) that

$$\mathbb{E}\left[T(s_1, s_2)\right] \leq \left(1 + \frac{1}{\sqrt{2}}\right) H^2(s, s_1) - \left(1 - \frac{1}{\sqrt{2}}\right) H^2(s, s_2)$$

and thus,

$$\begin{aligned} \mathbb{P}\left[T(s_1, s_2) \geq x\right] &= \mathbb{P}\left[Z(s_1, s_2) \geq x - \mathbb{E}\left[T(s_1, s_2)\right]\right] \\ &\leq \mathbb{P}\left[Z(s_1, s_2) \geq x - \left(1 + \frac{1}{\sqrt{2}}\right) H^2(s, s_1) + \left(1 - \frac{1}{\sqrt{2}}\right) H^2(s, s_2)\right]. \end{aligned}$$

Moreover, the random variable $Z(s_1, s_2)$ can be written as

$$Z(s_1, s_2) = \frac{1}{n}\sum_{i=1}^{n}\left(Z_i(s_1, s_2) - \mathbb{E}\left[Z_i(s_1, s_2)\right]\right)$$

where for each $i \in \{1, \ldots n\}$,

$$Z_i(s_1, s_2) = \int_{\mathbb{T}} \psi\left(s_1(\cdot, x_i), s_2(\cdot, x_i)\right) \mathrm{d}N_i - \mathbb{E}\left(\int_{\mathbb{T}} \psi\left(s_1(\cdot, x_i), s_2(\cdot, x_i)\right) \mathrm{d}N_i\right).$$

We apply Lemma 6.3 with $f_i = \psi\left(s_1(\cdot, x_i), s_2(\cdot, x_i)\right)$, and $\upsilon = H^2(s, s_1) + H^2(s, s_2) + H^2(s_1, s_2)$. The calculus of υ is provided by the proof of Proposition 3 of Baraud (2010). Consequently, if $r = x - \left(1 + \frac{1}{\sqrt{2}}\right) H^2(s, s_1) + \left(1 - \frac{1}{\sqrt{2}}\right) H^2(s, s_2)$ is non-negative,

$$\mathbb{P}\left[T(s_1, s_2) \geq x\right] \leq \exp\left(-\frac{nr^2}{2\upsilon + \frac{r\sqrt{2}}{3}}\right).$$

To avoid some complicated calculus, we bound by above the preceding probability without optimisation and we do not explicit the function a. We merely say that $r \geq x + CH^2(s_1, s_2)$ for $C = -(1 + 1/\sqrt{2})\kappa^{-2} + \frac{1-1/\sqrt{2}}{2(1+\kappa^{-2})}$. For κ large enough, $C > 0$ and hence,

$$\begin{aligned} \mathbb{P}\left[T(s_1, s_2) \geq x\right] &\leq \exp\left(-\frac{n\left(x + CH^2(s_1, s_2)\right)^2}{6(1+C)H^2(s_1, s_2) + \frac{\sqrt{2}}{3}\left(x + CH^2(s_1, s_2)\right)}\right) \\ &\leq \exp\left[-an\left(x + CH^2(s_1, s_2)\right)\right] \end{aligned}$$

where a is some positive function of C. If $x + CH^2(s_1, s_2) \geq 0$ then $r \geq 0$ and thus the preceding inequality holds. If $x+CH^2(s_1, s_2) < 0$ this inequality still holds since $\mathbb{P}\left[T(s_1, s_2) \geq x\right] \leq 1$. $\square$

6.1.3. *Proof of Proposition 2.1.* We define $\mathbb{V} = \cup_{\mathscr{F}\in\mathfrak{F}} \mathbb{V}_{\mathscr{F}}$ and $\bar{\Delta}$ on $\mathbb{V}$ by

$$\bar{\Delta}(V) = \inf_{\mathscr{F}\in\mathfrak{F}, \mathbb{V}_{\mathscr{F}}\ni V} \left(\Delta_{\mathscr{F}}(V) + \Delta\left(\mathscr{F}\right)\right).$$

The assumptions of Theorem 2.1 are satisfied for $(\mathbb{V}, \bar{\Delta})$ since

$$\sum_{V\in\mathbb{V}} e^{-\bar{\Delta}(V)} \quad \leq \quad \sum_{V\in\mathbb{V}}\sum_{\mathscr{F}\in\mathfrak{F}} e^{-\Delta_{\mathscr{F}}(V)-\Delta(\mathscr{F})} \mathbb{1}_{V\in\mathbb{V}_{\mathscr{F}}} \leq \sum_{\mathscr{F}\in\mathfrak{F}}\sum_{V\in\mathbb{V}_{\mathscr{F}}} e^{-\Delta_{\mathscr{F}}(V)-\Delta(\mathscr{F})} \leq 1.$$

Consequently, there exists an estimator $\hat{s}$ such that:

$$C\mathbb{E}\left[H^2(s,\hat{s})\right] \leq \inf_{\mathscr{F}\in\mathfrak{F}} \left(\inf_{f\in\mathscr{F}} \left\{d_2^2(\sqrt{s}, f) + \varepsilon_{\mathscr{F}}(f)\right\} + \frac{\Delta(\mathscr{F})}{n}\right)$$

with $\varepsilon_{\mathscr{F}}(f) = \inf_{V\in\mathbb{V}_{\mathscr{F}}} \left\{d_2^2(f, V) + \eta_V^2 + \frac{\Delta_{\mathscr{F}}(V)}{n}\right\}$ and where C is an universal positive constant. $\square$

6.1.4. *Proof of Theorem 2.3.* For any function f on $\mathbb{T}\times\mathcal{X}$, we set $\bar{f}$ the function on $\mathbb{T}\times\mathbb{X}$ defined by $\bar{f}(t,x) = f(t, x(t))$. Let for any $V\in\mathbb{V}$, $\bar{V} = \left\{\bar{f},\, f\in V\right\}$ and let $\bar{\mathbb{V}} = \{\bar{V},\, V\in\mathbb{V}\}$. We define $\bar{\Delta}$ on $\bar{\mathbb{V}}$ by $\bar{\Delta}(\bar{V}) = \Delta(V)$. Theorem 2.1 applied with $(\bar{V}, \bar{\Delta})$ leads to an estimator $\hat{s}$ such that, for all $f\in\mathbb{T}\times\mathcal{X}$,

$$C\mathbb{E}\left[H^2(s,\hat{s})\right] \leq d_2^2\left(\sqrt{s}, \bar{f}\right) + \inf_{V\in\mathbb{V}} \left\{d_2^2\left(\bar{f}, \bar{V}\right) + \frac{\dim \bar{V} + \Delta(V)}{n}\right\}.$$

The proof follows from the fact that $\dim\bar{V} \leq \dim V$ and that $d_2^2\left(\bar{f}, \bar{V}\right) \leq d_\infty^2\left(f, V\right)$. $\square$

6.2. **Proofs of Section 3.**

6.2.1. *Proof of Corollary 3.1.* This corollary can be easily deduced from Lemma 1 and Proposition 1 of Baraud and Birgé (2011). $\square$

6.2.2. *Proof of Corollary 3.2.* For all $j\in\{1,\ldots,k_1\}$ we consider the function $u_j(t,x) = t$, and for $j\in\{k_1+1,\ldots,k_1+l\}$, the function $u_j(t,x) =< \theta, x >$. The function f can be written as the composite function

$$f = g\circ(u_1,\ldots,u_{k_1},u_{k_1+1},\ldots,u_{k_1+l}).$$

Let $\mathbb{W}$ be a collection of finite dimensional linear spaces of bounded functions on $[0,1]^{k_1}\times[-1,1]^l$ and $\Delta\geq 1$ be a map on $\mathbb{W}$ such that $\sum_{W\in\mathbb{W}} e^{-\Delta(W)} \leq 1$. We apply Corollary 1 of Baraud and Birgé (2011) with $\mathbb{T}_j = \{\{(t,x)\mapsto t\}\}$ for $j\in\{1,\ldots,k_1\}$,

$\mathbb{T}_j = \left\{\{(t,x) \mapsto < x, \theta >, \theta \in \mathbb{R}^{k_2}\}\right\}$ for $j \in \{k_1+1, \ldots, k_1+l\}$ and $\mathbb{F} = \mathbb{W}$. This leads to an estimator $\hat{s}$ such that, for all $W \in \mathbb{W}$,

$$\begin{aligned} C\mathbb{E}\left[H^2(s,\hat{s})\right] &\leq d_2^2\left(\sqrt{s}, f\right) + \frac{\ln n \vee \ln\left(\|g\|_{\boldsymbol{\alpha}}^2 k_2^{-1}\right) \vee 1}{n} k_2 \\ &\quad + d_\infty^2(g, W) + \frac{\dim W + \Delta(W)}{n}, \end{aligned}$$

where $C > 0$ depends on $\boldsymbol{\alpha}$, k_1 and l. We finally choose for $\mathbb{W}$ the collection of linear spaces given by Proposition 1 of Baraud and Birgé (2011) and use their Lemma 1.

6.2.3. *Proof of Corollary 3.3.* Let $v \in \mathcal{H}^{\boldsymbol{\beta}}\left([0,1]^{k_2}\right)$ satisfying to $v^2 \in \mathcal{H}^{\gamma}\left([0,1]^{k_2}\right)$, let $g \in \mathcal{H}\left([0, \|v\|_\infty^2]\right)$ and let $f \in \mathscr{F}$ be the function $f(t,x) = v(x)g(tv^2(x))$. Let us consider the functions on $\mathbb{T} \times \mathbb{X}$

$$v_1(t,x) = v(x)/\|v\|_\infty, \quad v_2(t) = t, \quad v_3(t,x) = v^2(t)/\|v\|_\infty^2,$$

and let φ be a function on $[-1,1]^3$ defined by

$$\forall a,b,c \in [-1,1], \quad \varphi(a,b,c) = \|v\|_\infty a g\left(\|v\|_\infty^2 bc\right).$$

The function f can be thus written as the composite function $\varphi \circ (v_1, v_2, v_3)$.

Furthermore, according to Definition 1 of Baraud and Birgé (2011), we can choose for the modulus of continuity of φ the function w_φ defined by

$$\forall x,y,z \in [0,2], \quad \mathrm{w}_\varphi(x,y,z) = \|v\|_\infty \left(\|g\|_\infty x, \|v\|_\infty^{2(\alpha\wedge 1)}\|g\|_\alpha y^{\alpha\wedge 1}, \|v\|_\infty^{2(\alpha\wedge 1)}\|g\|_\alpha z^{\alpha\wedge 1}\right).$$

Let then $\mathbb{V}_1$ and $\mathbb{V}_2$ be two collections of finite dimensional linear spaces of $\mathbb{L}^2(\mathbb{X}, \nu_n)$ to approximate v and v^2 respectively. Let us note $\bar{g}(x) = \|v\|_\infty g(\|v\|_\infty^2 x)$ and let $\bar{\mathbb{W}}$ be a collection of finite dimensional linear spaces of $\mathbb{L}^\infty([0,1])$ to approximate $\bar{g}$. For each $\bar{W} \in \bar{\mathbb{W}}$, let $\bar{W}' = \left\{(a,b,c) \mapsto a\bar{h}(bc), \bar{h} \in \bar{W}\right\}$ and let $\bar{\mathbb{W}}'$ the collection of all the $\bar{W}'$ when $\bar{W}$ varies among $\bar{\mathbb{W}}$.

We apply Theorem 2 of Baraud and Birgé (2011) with $\mathbb{T}_1 = \mathbb{V}_1$, $\mathbb{T}_2 = \{\{(t,x) \mapsto t\}\}$, $\mathbb{T}_3 = \mathbb{V}_2$ and $\mathbb{F} = \bar{\mathbb{W}}'$. This leads to an estimator $\hat{s}$ such that, for all $V_1 \in \mathbb{V}_1$, $V_2 \in \mathbb{V}_2$ and $\bar{W} \in \bar{\mathbb{W}}$,

$$\begin{aligned} C\mathbb{E}\left[H^2(s,\hat{s})\right] &\leq d_2^2\left(\sqrt{s}, f\right) + \|g\|_\infty^2 d_{\mathtt{x}}^2(v, V_1) + \frac{\ln n \vee \ln(\|v\|_\infty \|g\|_\infty) \vee 1}{n} (\dim V_1 \vee 1) \\ &\quad + \|v\|_\infty^2 \|g\|_\alpha^2 d_{\mathtt{x}}^{2(\alpha\wedge 1)}(v^2, V_2) + \frac{\ln n \vee \ln\left(\|v\|_\infty^{1+2(\alpha\wedge 1)}\|g\|_\alpha\right) \vee 1}{n} (\dim V_2 \vee 1) \\ &\quad + d_\infty^2(\bar{g}, \bar{W}) + \frac{\dim \bar{W} \vee 1}{n}. \end{aligned}$$

We conclude by choosing for $\mathbb{V}_1$, $\mathbb{V}_2$ and $\bar{\mathbb{W}}$ collections of piecewise constant polynomial linear spaces as in the preceding proofs. $\square$

6.2.4. *Proof of Lemma 3.1.* The first part of the lemma is given by Proposition 4 of Baraud and Birgé (2011). For the second part, we define g on $[0,1/2]$ by $g(x) = (1-2x)^\alpha$ and v on $[0,1]$ by $v(x) = \sqrt{1-1/2x^\beta}$. Then, $v \in \mathcal{H}^\beta([0,1])$ and $g \in \mathcal{H}^\alpha\left([0,\|v\|_\infty^2]\right)$. If $\alpha' > \alpha$, the function $f(t,0) = g(t)$ does not belong to $\mathcal{H}^{\alpha'}([0,1])$, whereas if $\alpha'\beta' > \alpha\beta$, the function $f(1/2,x) = 2^{-\alpha}\sqrt{1-1/2x^\beta}x^{\alpha\beta}$ does not belong to $\mathcal{H}^{\alpha'\beta'}([0,1])$. $\square$

6.3. **Proofs of Section 4.**

6.3.1. *Proof of Lemma 4.1.* The proof of this proposition requires the following elementary lemma.

Lemma 6.4. *Let $f, f' \in \mathbb{L}^2(\mathbb{T},\mu)$ and $g, g' \in \mathbb{L}^2(\mathbb{X},\nu_n)$ such that $\|f\|_{\mathtt{t}} = \|f'\|_{\mathtt{t}} = 1$ and $\|g\|_{\mathtt{x}} = \|g'\|_{\mathtt{x}} = 1$. Let $\kappa, \kappa' \in \mathbb{R}$. The following inequality holds:*

$$d_2^2\left(\kappa f g, \kappa' f' g'\right) = \left(\kappa - \kappa'\right)^2 + \kappa\kappa'\left(d_{\mathtt{t}}^2(f,f') + d_{\mathtt{x}}^2(g,g') - 1/2\, d_{\mathtt{t}}^2(f,f') d_{\mathtt{x}}^2(g,g')\right).$$

In this proof, we say that a set $S(\eta)$ is a η-net of an other set V in a metric space (E,d) if, for all $y \in V$, there exists $x \in S(\eta)$ such that $d(x,y) \leq \eta$.

Let us denote by S_1 (respectively S_2) the unit sphere of V_1 (respectively V_2). Let now for any $\eta > 0$, $S_1(\eta) \subset S_1$ (respectively $S_2(\eta) \subset S_2$) be a η-net of S_1 (respectively S_2) such that:

$$\begin{aligned} \forall (f,g) \in \mathbb{L}^2(\mathbb{T},\mu) \times \mathbb{L}^2(\mathbb{X},\nu_n),\ \forall x \geq 0, \quad |S_1(\eta) \cap \mathcal{B}_t(f, x\eta)| &\leq (2x+1)^{\dim V_1} \\ |S_2(\eta) \cap \mathcal{B}_{\mathtt{x}}(g, x\eta)| &\leq (2x+1)^{\dim V_2} \end{aligned}$$

where $\mathcal{B}_t(f,x\eta)$ and $\mathcal{B}_{\mathtt{x}}(g,x\eta)$ are balls centered at f and g with radius $x\eta$ of the metric spaces $(\mathbb{L}^2(\mathbb{T},\mu), d_{\mathtt{t}})$ and $(\mathbb{L}^2(\mathbb{X},\nu_n), d_{\mathtt{x}})$ respectively. We refer to Lemma 4 of Birgé (2006) for the existence of these nets. Let now

$$S(\eta) = \bigcup_{k \in \mathbb{N}^\star} \left\{ \frac{k\eta}{\sqrt{2}} f g,\ (f,g) \in S_1\left(\frac{1}{\sqrt{2}k}\right) \times S_2\left(\frac{1}{\sqrt{2}k}\right) \right\}.$$

First of all, $S(\eta)$ is a η-net of V. Indeed, let $\varphi \in V$. We can write $\varphi(t,x) = \kappa f(t) g(x)$ with $\kappa \in \mathbb{R}_+$, $f \in S_1$ and $g \in S_2$. Let us define $k = \inf\left\{i \in \mathbb{N},\ i \geq \sqrt{2}\kappa_1/\eta\right\}$ and let $(f', g') \in S_1(1/(\sqrt{2}k)) \times S_2(1/(\sqrt{2}k))$ such that

$$d_{\mathtt{t}}(f,f') \leq \frac{1}{\sqrt{2}k} \quad \text{and} \quad d_{\mathtt{x}}(g,g') \leq \frac{1}{\sqrt{2}k}.$$

Then, by Lemma 6.4, the application $\varphi'(t,x) = \frac{k\eta}{\sqrt{2}} f'(t) g'(x)$ is such that $d_2(\varphi,\varphi') \leq \eta$.

Furthermore, let φ be a function of $S(\eta)$ written as $\varphi(t,x) = \kappa f(t) g(x)$. We want to bound the cardinality of the set $S(\eta) \cap \mathcal{B}(\kappa f g, x\eta)$ (where $\mathcal{B}(\kappa f g, x\eta)$ is the ball centered at $\kappa f g$ with radius $x\eta$ of the metric space $(\mathbb{L}^2(\mathbb{T}\times\mathbb{X}, M), d_2)$). Let then $\varphi' = \kappa' f' g' \in S(\eta) \cap \mathcal{B}(\kappa f g, x\eta)$. We derive from Lemma 6.4 that:

$$(\kappa - \kappa')^2 \leq d_2^2(\varphi,\varphi') \leq x^2\eta^2.$$

Consequently, κ' has to be chosen among at most $2\sqrt{2}x+1$ numbers. Afterwards, assume that

$$\left(\int_{\mathbb{T}} f(t)f'(t)\,\mathrm{d}\mu(t)\right)\left(\int_{\mathbb{X}} g(x)g'(x)\,\mathrm{d}\nu_n(x)\right) < 0.$$

In this situation, $d_2^2(\varphi,\varphi') \geq \kappa^2+\kappa'^2$. Since, $\kappa \geq \eta/\sqrt{2}$ and $\kappa' \leq \kappa + x\eta$, for any $x \geq 2$, $\frac{\kappa'}{\kappa} \leq 1+\sqrt{2}x \leq \frac{3}{2}x$. Thus, $d_2^2(\varphi,\varphi') \geq \frac{4\kappa'^2}{9x^2}$, and since

$$\|f-f'\|_{\mathtt{t}}^2 \leq 4 \quad \text{and} \quad \|g-g'\|_{\mathtt{x}}^2 \leq 4,$$

we have,

$$f' \in S_1\left(\frac{\eta}{2\kappa'}\right) \cap \mathcal{B}_{\mathtt{t}}\left(f, 6x^2\frac{\eta}{2\kappa'}\right) \quad \text{and} \quad g' \in S_2\left(\frac{\eta}{2\kappa'}\right) \cap \mathcal{B}_{\mathtt{x}}\left(g, 6x^2\frac{\eta}{2\kappa'}\right).$$

This permits us to control the number of f' and g' possible. If now,

$$\int_{\mathbb{T}} f(t)f'(t)\,\mathrm{d}\mu(t) > 0 \quad \text{and} \quad \int_{\mathbb{X}} g(x)g'(x)\,\mathrm{d}\nu_n(x) > 0,$$

we derive from Lemma 6.4 and from the elementary inequality $x+y-1/2xy \geq 1/2(x+y)$ for $x,y \in [0,2]$, that:

$$(\kappa-\kappa')^2 + \frac{\kappa\kappa'}{2}\left(d_{\mathtt{t}}^2(f,f') + d_{\mathtt{x}}^2(g,g')\right) \leq d_2^2\left(\varphi,\varphi'\right) \leq x^2\eta^2,$$

which leads to

$$d_{\mathtt{t}}^2(f,f') + d_{\mathtt{x}}^2(g,g') \leq \frac{2x^2\eta^2}{\kappa\kappa'} \leq \frac{3x^3\eta^2}{\kappa'^2}.$$

Thus,

$$f' \in S_1\left(\frac{\eta}{2\kappa'}\right) \cap \mathcal{B}_{\mathtt{t}}\left(f, 2\sqrt{3}x^{3/2}\frac{\eta}{2\kappa'}\right) \quad \text{and} \quad g' \in S_2\left(\frac{\eta}{2\kappa'}\right) \cap \mathcal{B}_{\mathtt{x}}\left(g, 2\sqrt{3}x^{3/2}\frac{\eta}{2\kappa'}\right).$$

Finally, if

$$\int_{\mathbb{T}} f(t)f'(t)\,\mathrm{d}\mu(t) < 0 \quad \text{and} \quad \int_{\mathbb{X}} g(x)g'(x)\,\mathrm{d}\nu_n(x) < 0,$$

it comes from the preceding calculus that

$$f' \in S_1\left(\frac{\eta}{2\kappa'}\right) \cap \mathcal{B}_{\mathtt{t}}\left(-f, 2\sqrt{3}x^{3/2}\frac{\eta}{2\kappa'}\right) \quad \text{and} \quad g' \in S_2\left(\frac{\eta}{2\kappa'}\right) \cap \mathcal{B}_{\mathtt{x}}\left(-g, 2\sqrt{3}x^{3/2}\frac{\eta}{2\kappa'}\right).$$

Consequently, we have proved

$$\forall \varphi \in S(\eta),\ \forall x \geq 2, \quad |S(\eta) \cap \mathcal{B}(\varphi, x\eta)| \leq \left(2\sqrt{2}x+1\right)\left(12x^2+1\right)^{\dim V_1 + \dim V_2}.$$

To prove the result, this inequality must hold for all $\varphi \in \mathbb{L}^2(\mathbb{T}\times\mathbb{X}, M)$. If $\varphi \in \mathbb{L}^2(\mathbb{T}\times\mathbb{X}, M)$, may be $|S(\eta) \cap \mathcal{B}(\varphi, x\eta)| = 0$. If not, there exists $\varphi' \in S(\eta) \cap \mathcal{B}(\varphi, x\eta)$ and thus,

$$|S(\eta) \cap \mathcal{B}(\varphi, x\eta)| \leq \left|S(\eta) \cap \mathcal{B}\left(\varphi', 2x\eta\right)\right|.$$

Consequently,

$$\forall \varphi \in \mathbb{L}^2(\mathbb{T}\times\mathbb{X}, M),\ \forall x \geq 2, \quad |S(\eta) \cap \mathcal{B}(\varphi, x\eta)| \leq \left(4\sqrt{2}x+1\right)\left(48x^2+1\right)^{\dim V_1 + \dim V_2}.$$

The conclusion ensues from the elementary inequality

$$\forall x \geq 2, \quad 48x^2+1 \leq e^{1.4x^2}.$$

$\square$

6.3.2. *Proof of Proposition 4.1.* For any pair $(V_1, V_2) \in \mathbb{V}_1 \times \mathbb{V}_2$, we define the set V by relation (10). Let then $\mathbb{V}$ be the collection of all V when (V_1, V_2) varies among $\mathbb{V}_1 \times \mathbb{V}_2$. Let $\bar{\Delta}$ be the application on $\mathbb{V}$ defined by

$$\bar{\Delta}\,(V) = \Delta_1(V_1) + \Delta_2(V_2)$$

when V corresponds to (V_1, V_2). We apply afterwards Theorem 2.1 with $\mathbb{V}$ and $\bar{\Delta}$ to derive an estimator $\hat{s}$ such that

$$C\mathbb{E}\left[H^2(s,\hat{s})\right] \leq \inf_{V \in \mathbb{V}} \left\{ d_2^2\left(\sqrt{s}, V\right) + \frac{\dim V_1 + \dim V_2 + \Delta_1(V_1) + \Delta_2(V_2)}{n} \right\}.$$

Let thus $\kappa v_1 v_2 \in \mathscr{F}$, and let $(v_1', v_2') \in V_1 \times V_2$ such that $\|v_1'\|_{\mathtt{t}} = \|v_2'\|_{\mathtt{x}} = 1$. The preceding inequality implies

$$C'\mathbb{E}\left[H^2(s,\hat{s})\right] \leq d_2^2\left(\sqrt{s}, \kappa v_1 v_2\right) + \kappa^2 d_2^2\left(v_1 v_2, v_1' v_2'\right) + \frac{\dim V_1 + \dim V_2 + \Delta_1(V_1) + \Delta_2(V_2)}{n}.$$

Some calculus shows that

$$d_2^2\left(v_1 v_2, v_1' v_2'\right) \leq 2\left(d_{\mathtt{t}}^2\left(v_1, v_1'\right) + d_{\mathtt{x}}^2\left(v_2, v_2'\right)\right).$$

By taking the infimum over all v_1' and v_2',

$$\begin{aligned} C''\mathbb{E}\left[H^2(s,\hat{s})\right] \quad \leq \quad & d_2^2\left(\sqrt{s}, \kappa v_1 v_2\right) + \kappa^2 d_{\mathtt{t}}^2\left(v_1, S_1\right) + \frac{\dim V_1 + \Delta_1(V_1)}{n} + \kappa^2 d_{\mathtt{x}}^2\left(v_2, S_2\right) \\ & + \frac{\dim V_2 + \Delta_2(V_2)}{n} \end{aligned}$$

where S_1 and S_2 are the unit spheres of V_1 and V_2 respectively. We conclude by using the fact that $d_{\mathtt{t}}(v_1, S_1) \leq 2 d_{\mathtt{t}}(v_1, V_1)$, $d_2(v_2, S_2) \leq 2 d_{\mathtt{x}}(v_2, V_2)$ and by taking the infimum over all $(V_1, V_2) \in \mathbb{V}_1 \times \mathbb{V}_2$. $\square$

6.3.3. *Proof of Lemma 4.2.* We derive from some calculus that for all $-1/2 < r \leq b, b' \leq R$,

$$\int_0^1 \left(\sqrt{2b+1}t^b - \sqrt{2b'+1}t^{b'}\right)^2 \mathrm{d}t = \frac{4\left(b-b'\right)^2}{(1+b+b')\left(\sqrt{2b+1} + \sqrt{2b'+1}\right)^2}$$

and thus

$$\frac{(b-b')^2}{(1+2R)^2} \leq \int_0^1 \left(\sqrt{2b+1}t^b - \sqrt{2b'+1}t^{b'}\right)^2 \mathrm{d}t \leq \frac{(b-b')^2}{(1+2r)^2}.$$

This concludes the first part of the proof. For the second part, for $b > 0$, we define

$$g_b(t) = \frac{f_b(t)}{\|f_b\|_{\mathtt{t}}} = \frac{2^{k/2} b^{1/2+k/2}}{\sqrt{k/2\,k!}}\, t^{k/2} e^{-bt}.$$

Then for $0 < r \leq b, b' \leq R$,

$$\begin{aligned}
\frac{1}{2}\int_0^\infty \left(g_b(t) - g_{b'}(t)\right)^2 \,\mathrm{d}t &= 1 - \frac{\left(2\sqrt{bb'}\right)^{k+1}}{(b+b')^{k+1}} \\
&= \frac{\sum_{j=0}^{k}(b+b')^j(2\sqrt{bb'})^{k-j}}{(b+b')^{k+1}(\sqrt{b}+\sqrt{b'})^2}\left(b-b'\right)^2 \\
&= \frac{1}{(b+b')(\sqrt{b}+\sqrt{b'})^2}\sum_{j=0}^{k}\left(\frac{2\sqrt{bb'}}{b+b'}\right)^j\left(b-b'\right)^2.
\end{aligned}$$

Consequently the lemma ensues from the inequality below.

$$\frac{1}{8R^2}\left(b-b'\right)^2 \leq \frac{1}{2}\int_0^\infty \left(g_b(t) - g_{b'}(t)\right)^2 \,\mathrm{d}t \quad \leq \quad \frac{k+1}{8r^2}\left(b-b'\right)^2.$$

6.3.4. *Proof of Proposition 4.3.* We generalize Lemma 4.1 for some new spaces. The proof of the following lemma is analogue to the one of Lemma 4.1 and will not be detailed.

Lemma 6.5. *Let V_1 and V_2 be some subsets of the unit sphere of $\mathbb{L}^2(\mathbb{T},\mu)$ and $\mathbb{L}^2(\mathbb{X},\nu_n)$ respectively. For each $i \in \{1,2\}$, we assume that there exist $W_\mathtt{i}$ a subset of a $D_{\bar{W}_\mathtt{i}}$-dimensional normed linear space $(\bar{W}_\mathtt{i}, |\cdot|_\mathtt{i})$, and a map Φ_i from $W_\mathtt{i}$ onto V_i such that:*

$$\forall (x,y) \in W_1, \quad \underline{\rho_1}|x-y|_1 \leq d_\mathtt{t}(\Phi_1(x),\Phi_1(y)) \leq \bar{\rho}_1|x-y|_1 \tag{19}$$

$$\forall (x,y) \in W_2, \quad \underline{\rho_2}|x-y|_2 \leq d_\mathtt{x}(\Phi_2(x),\Phi_2(y)) \leq \bar{\rho}_2|x-y|_2. \tag{20}$$

The set

$$V = \{\kappa v_1 v_2,\ (v_1,v_2) \in V_1 \times V_2,\ \kappa \geq 0\}$$

has a finite metric dimension bounded by

$$D_V = C\left[1 + D_{\bar{W}_1}\ln\left(1+\frac{\bar{\rho}_1}{\underline{\rho_1}}\right) + D_{\bar{W}_2}\ln\left(1+\frac{\bar{\rho}_2}{\underline{\rho_2}}\right)\right]$$

where C is an universal constant.

Lemma 6.6. *Let for all $b_0 < r \leq R$, $V_\mathtt{t}(r,R)$ be the set defined by*

$$V_\mathtt{t}(r,R) = \left\{\frac{u_b}{\|u_b\|_\mathtt{t}},\ r \leq b \leq R\right\}.$$

Condition (19) holds with $D_{\bar{W}_1} = 1$, $\underline{\rho_1} = \underline{\rho}(R)$ and $\bar{\rho_1} = \bar{\rho}(r)$.

Lemma 6.7. *For any integer $\rho > 0$ and $W \in \mathbb{W}$, let*

$$V_2(W,R) = \left\{\frac{v_\theta}{\|v_\theta\|_\mathtt{x}},\ \theta \in W,\ \|\theta\|_2 \leq \rho\right\}.$$

There exists $(\bar{W}_2, |\cdot|_2)$ a finite dimensional linear space and Φ_2 a map from $\bar{W}_2$ onto $V_2(W,R)$ such that condition (20) holds with $D_{\bar{W}_2} \leq \dim W$, $\underline{\rho_2} = e^{-3\rho}$ and $\bar{\rho}_2 = e^{3\rho}$.

Proof of Lemma 6.7. For any integers $i, j \in \mathbb{N}^\star$, let us denote by $\varphi_{i,j}$ the linear form on $\mathbb{R}^{k_2}$ defined by $\varphi_{i,j}(\theta) =< x_i - x_j, \theta >$. Let $W_1 = \cap_{i \neq j} \mathrm{Ker}\, \varphi_{i,j}$ and let W_2 such that $W = W_1 \oplus W_2$ and such that $< u, v >= 0$ for all $(u, v) \in W_1 \times W_2$. Since the functions of $\mathbb{L}^2(\mathbb{X}, \nu_n)$ are defined ν_n-almost everywhere, the set $V_2(W, R)$ can be written as

$$V_2(W, R) = \Phi_2\left(\{\theta \in W_2,\, \|\theta\| \leq \rho\}\right) \quad \text{where} \quad \Phi_2(\theta) = \frac{v_\theta}{\|v_\theta\|_{\mathtt{x}}}.$$

Let, for all $x \in \mathbb{X}$, Ψ_x be the function defined from $\mathbb{X}$ into $\mathbb{R}$ by $\Psi_x(\theta) = \Phi_2(\theta)(x)$. We derive from some calculus that the differential of Ψ_x at the point $\theta \in W_2$, denoted by $d\Psi_x(\theta)$, satisfies

$$\forall h \in \mathbb{R}^{k_2}, \quad d\Psi_x(\theta) \cdot h = \frac{\frac{1}{n}\sum_{i=1}^n \exp\left(2 < \theta, x_i > + < \theta, x >\right)\left(< x - x_i, h >\right)}{\left(\frac{1}{n}\sum_{i=1}^n \exp\left(2 < \theta, x_i >\right)\right)^{3/2}}.$$

In particular, we have

$$\forall h \in \mathbb{R}^{k_2}, \quad \frac{e^{-3\rho}}{n} \sum_{i=1}^n |< x - x_i, h >| \leq |d\Psi_x(\theta) \cdot h| \leq \frac{e^{3\rho}}{n} \sum_{i=1}^n |< x - x_i, h >|.$$

If we endow W_2 with the norm $|\cdot|_2$ defined by

$$\forall \theta \in W_2, \quad |\theta|_2 = \sqrt{\frac{1}{n}\sum_{i=1}^n \left(\frac{1}{n}\sum_{j=1}^n |< x_i - x_j, \theta >|\right)^2},$$

the mean value theorem leads to

$$\forall (\theta_1, \theta_2) \in W_2, \quad e^{-3\rho}|\theta_1 - \theta_2|_2 \leq d_{\mathtt{x}}\left(\Phi_2(\theta_1), \Phi_2(\theta_2)\right) \leq e^{3\rho}|\theta_1 - \theta_2|_2,$$

which concludes the proof. $\square$

Thanks to Lemma 6.5, for all $\rho > 0$, $c < r \leq R$ and $W \in \mathbb{W}$ the set

$$V(r, R, W, \varrho) = \{a u_b v_\theta,\, a \geq 0,\, r \leq b \leq R,\, \theta \in W,\, \|\theta\| \leq \varrho\}$$

has a metric dimension bounded by

$$CD_{V(r,R,W,\varrho)} = 1 + \varrho \dim W + \ln\left(1 + \frac{\bar{\rho}(r)}{\underline{\rho}(R)}\right)$$

for some universal positive constant C. Let us define the collection $\mathbb{V}$ by

$$\mathbb{V} = \left\{V\left(b_0 + 1/r, b_0 + R, W, \varrho\right),\, W \in \mathbb{W},\, r, R, \varrho \in \mathbb{N}^\star\right\}$$

and the map $\bar{\Delta}$ on $\mathbb{V}$ by

$$\bar{\Delta}\left(V(r, R, W, \varrho)\right) = \Delta(W) + \ln(2R^2) + \ln(2r^2) + \ln(2\varrho^2).$$

Theorem 2.1 applied with $(\mathbb{V}, \bar{\Delta})$ provides an estimator $\hat{s}$, such that, for all $W \in \mathbb{W}$, all $\varrho, r, R \in \mathbb{N}^\star$, all $\theta' \in W$ such that $\|\theta'\| \leq \varrho$, all $a \geq 0$, all b satisfying to $b_0 + 1/r \leq b \leq b_0 + R$,

$$C'\mathbb{E}\left[H^2(s, \hat{s})\right] \leq d_2^2(\sqrt{s}, a u_b v_{\theta'}) + \frac{1 + \varrho \dim W + \ln\left(1 + \frac{\bar{\rho}(b_0 + 1/r)}{\underline{\rho}(b_0 + R)}\right) + \ln r + \ln R + \ln \varrho}{n}$$

where C' is another universal positive constant. In particular, with $R = \inf\{i \in \mathbb{N}^\star, i \geq b - b_0\}$, $r = \inf\{i \in \mathbb{N}^\star, i \geq 1/(b-b_0)\}$, $\varrho = \inf\{i \in \mathbb{N}^\star, i \geq \|\theta'\|\}$ this inequality leads to

$$\begin{aligned} C''\mathbb{E}\left[H^2(s,\hat{s})\right] &\leq d_2^2(\sqrt{s}, au_b v_{\theta'}) + \frac{(1 \vee \|\theta'\|)(1 \vee \dim W)}{n} \\ &\quad + \frac{1}{n}\left\{\ln\left[1 \vee \bar{\rho}\left(b_0 + \frac{b-b_0}{b-b_0+1}\right)\right] + \left|\ln\left(1 \wedge \underline{\rho}(1+b)\right)\right| + \left|\ln(b-b_0)\right|\right\} \end{aligned}$$

where C'' is another universal positive constant. For $\theta \in \mathbb{R}^{k_2}$, the triangular inequality leads to

$$\begin{aligned} d_2^2(\sqrt{s}, au_b v_{\theta'}) &\leq 2\left(d_2^2(\sqrt{s}, au_b v_\theta) + d_2^2(au_b v_\theta, au_b v_{\theta'})\right) \\ &\leq 2\left(d_2^2(\sqrt{s}, au_b v_\theta) + a^2\|u_b\|_{\mathtt{t}}^2 d_{\mathtt{x}}^2(v_\theta, v_{\theta'})\right). \end{aligned}$$

Some calculus shows that $d_{\mathtt{x}}(v_\theta, v_{\theta'}) \leq e^{\|\theta\| \vee \|\theta'\|}\|\theta - \theta'\|$. We conclude by choosing θ' the projection of θ on W, and by taking the infimum over all $W \in \mathbb{W}$. □

6.4. **Proofs of Section 5.**

6.4.1. *Proofs of Lemmas 5.1 and 5.2.*

Proof of Lemma 5.1. We derive from some calculus that for $\theta_2, \theta_2' \in [-1/2+1/i_2, +\infty[$

$$\int_0^1 (t^{\theta_2} - t^{\theta_2'})^2 \,\mathrm{d}t = \frac{2(\theta_2 - \theta_2')^2}{(1+2\theta_2)(1+\theta_2+\theta_2')(1+2\theta_2')} \leq 2i_2^3(\theta_2 - \theta_2')^2.$$

Thus, for $(\theta_1, \theta_2), (\theta_1', \theta_2') \in K_{i_1,i_2}$, by the triangular inequality

$$\begin{aligned} \sqrt{\int_0^1 (\theta_1 t^{\theta_2} - \theta_1' t^{\theta_2'})^2 \,\mathrm{d}t} &\leq \frac{|\theta_1 - \theta_1'|}{\sqrt{2\theta_2'+1}} + \frac{\sqrt{2}\theta_1|\theta_2 - \theta_2'|}{\sqrt{(1+2\theta_2)(1+\theta_2+\theta_2')(1+2\theta_2')}} \\ &\leq i_2^{1/2}|\theta_1 - \theta_1'| + \sqrt{2}i_1 i_2^{3/2}|\theta_2 - \theta_2'|. \end{aligned}$$

□

Proof of Lemma 5.2. We derive from some calculus that for $\theta_2, \theta_2' \geq 1/i_2$, and $g(x) = x^{-k}$,

$$\int_0^\infty \left(\sqrt{t^{k-1}e^{-\theta_2 t}} - \sqrt{t^{k-1}e^{-\theta_2' t}}\right)^2 \,\mathrm{d}t = (k-1)!\left(g(\theta_2) + g(\theta_2') - 2g\left(\frac{\theta_2+\theta_2'}{2}\right)\right).$$

Consequently, by the mean value theorem,

$$\int_0^\infty \left(\sqrt{t^{k-1}e^{-\theta_2 t}} - \sqrt{t^{k-1}e^{-\theta_2' t}}\right)^2 \,\mathrm{d}t \leq (k-1)!\left(\sup_{c \in [\theta_2, \theta_2']} |g'(c)|\right)\left|\theta_2 - \theta_2'\right|.$$

The conclusion follows from the triangular inequality as in the proof of the preceding lemma. □

6.4.2. *Proof of Theorem 5.1.* We start with the following proposition.

Proposition 6.1. *Assume that Assumption 5.1 holds and let $K \in \mathcal{K}$. Let for any $1 \leq j \leq k$, W_j be a linear subspace of $\mathbb{L}^2(\mathbb{X}, \nu_n)$ with finite dimension and Z_j be a bounded subset of W_j. Let then $\boldsymbol{\rho} \in (\mathbb{R}_+)^k$ such that for all $j \in \{1, \dots, k\}$, $Z_j \subset \mathcal{B}_{\mathtt{x}}(0, \rho_j) = \{g \in \mathbb{L}^2(\mathbb{X}, \nu_n),\, \|g\|_{\mathtt{x}} \leq \rho_j\}$. Let π_K be the projection on K in $\mathbb{R}^k$ for the distance d_r defined by*

$$\forall \theta, \theta' \in \mathbb{R}^k, \quad d_r(\theta, \theta') = \sum_{j=1}^{k} R_j(K) \left|\theta_j - \theta'_j\right|^{\alpha_j(K)}.$$

The set

$$V = \left\{ (t, x) \mapsto f_{\pi_K(u(x))}(t),\, u \in \prod_{j=1}^{k} Z_j \right\}$$

has a metric dimension bounded by

$$D_V(\eta) = \frac{1}{2} \vee \frac{1}{4} \sum_{j=1}^{k} \ln\left(1 + 2\left(\frac{kR_j(K)}{\eta}\right)^{1/\alpha_j(K)} \rho_j\right) \dim(W_j).$$

Remark: K is a closed set for the usual topology of $\mathbb{R}^k$. It is straightforward to verify that K is still a closed set for the metric space $(\mathbb{R}^k, d_r)$. This shows that the projection on the closed convex set K always exists.

Proof of Proposition 6.1. As in the proof of Lemma 4.1, we say that a set $S(\eta)$ is a η-net of an other set V in a metric space (E, d) if, for all $y \in V$, there exists $x \in S(\eta)$ such that $d(x, y) \leq \eta$. For the sake of simplicity, we shall denote all along this proof, $R_j = R_j(K)$, $\alpha_j = \alpha_j(K)$, and $\pi = \pi_K$.

Let for all $j \in \{1, \dots, k\}$, $\eta_j > 0$, and $Z'_j(\eta_j)$ be a maximal subset of Z_j such that $d_{\mathtt{x}}(x, y) > \eta_j$ for all $x \neq y \in Z'_j(\eta)$. This is a η_j-net of Z_j such that, by Lemma 4 of Birgé (2006),

$$|Z'_j(\eta_j)| \leq \left|Z'_j(\eta_j) \cap \mathcal{B}_{\mathtt{x}}(0, \rho_j)\right| \leq \left(\frac{2\rho_j}{\eta} + 1\right)^{\dim W_j}.$$

Remark that the set

$$S(\eta) = \left\{ (t, x) \mapsto f_{\pi(u(x))}(t),\, u \in \prod_{j=1}^{k} Z'_j\left(\left(\frac{\eta}{kR_j}\right)^{1/\alpha_j}\right) \right\}$$

is a η-net of V. Indeed, let $f \in V$ be the function of the form $f(t, x) = f_{\pi(u(x))}(t)$ and for any $1 \leq j \leq k$, let $v_j \in Z'_j((\eta/(kR_j))^{1/\alpha_j})$ such that $\|u_j - v_j\|_{\mathtt{x}} \leq (\eta/(kR_j))^{1/\alpha_j}$.

We define $v = (v_1, \dots, v_k)$ and $g \in S(\eta)$ by $g(t,x) = f_{\pi(v(x))}(t)$. Then,

$$\begin{aligned}
\|f - g\|_2^2 &= \frac{1}{n}\sum_{i=1}^n \left\| f_{\pi(u(x_i))}(\cdot) - f_{\pi(v(x_i))}(\cdot)\right\|_{\mathtt{t}}^2 \\
&\leq \frac{1}{n}\sum_{i=1}^n d_r^2\left(u(x_i), v(x_i)\right) \\
&\leq \frac{1}{n}\sum_{i=1}^n k \sum_{j=1}^k R_j^2 \left|u_j(x_i) - v_j(x_i)\right|^{2\alpha_j} \\
&\leq \eta^2.
\end{aligned}$$

Moreover for all $x \geq 2$ and $\varphi \in \mathbb{L}^2(\mathbb{T} \times \mathbb{X}, M)$,

$$\begin{aligned}
|S(\eta) \cap \mathcal{B}(\varphi, x\eta)| &\leq \prod_{j=1}^k \left| Z'_j\left(\left(\frac{\eta}{kR_j}\right)^{1/\alpha_j}\right)\right| \\
&\leq \prod_{j=1}^k \left(2\left(\frac{kR_j}{\eta}\right)^{1/\alpha_j}\rho_j + 1\right)^{\dim(W_j)} \\
&\leq \exp\left(\frac{1}{4}\sum_{j=1}^k \dim(W_j)\ln\left(2\left(\frac{kR_j}{\eta}\right)^{1/\alpha_j}\rho_j + 1\right)x^2\right)
\end{aligned}$$

which leads to the result. □

Lemma 6.8. *Assume that there exist $k \in \mathbb{N}^\star$, $a, b \in (\mathbb{R}_+)^k$ with $\max_{1\leq i\leq k} a_i \geq 1$, and $\min_{1\leq i\leq k} b_i \geq 1$ such that*

$$\forall \eta > 0, \quad D_V(\eta) \leq \sum_{i=1}^k a_i \ln\left(1 + \frac{b_i}{\eta}\right).$$

Then, there exists an universal positive constant C such that

$$C\eta_V^2 \leq \frac{\sum_{i=1}^k a_i \ln(1+b_i)}{n} + \frac{\sum_{i=1}^k a_i}{n}\ln\left(1 + \frac{n}{\sum_{i=1}^k a_i}\right).$$

Proof of Lemma 6.8. If one increases D_V, η_V will increase. Consequently, without lost of generality we can assume that:

$$D_V(\eta) = \begin{cases} 2\sum_{i=1}^k a_i \ln(2b_i) + 2\left(\sum_{i=1}^k a_i\right)\ln\left(\frac{1}{\eta}\right) & \text{if } \eta < 1 \\ 2\sum_{i=1}^k a_i \ln(2b_i) & \text{otherwise.}\end{cases}$$

Remark that for all $\alpha, \beta, y > 0$, the equation

$$\alpha + \beta \ln x = \frac{y}{2x^2}$$

has only one solution x given by

$$x^2 = \frac{y}{\beta W\left(\frac{e^{\frac{2\alpha}{\beta}}}{\beta} y\right)},$$

where W is the Lambert function, defined as the inverse function of $t \mapsto te^t$. Consequently, by setting

$$\alpha = \sum_{i=1}^k a_i \ln(2b_i) \quad \text{and} \quad \beta = \sum_{i=1}^k a_i$$

we derive that the positive number η defined by

$$\eta^2 = \begin{cases} \frac{\beta W\left(\frac{e^{\frac{2\alpha}{\beta}}}{\beta} n\right)}{n} & \text{if } n > 2\alpha \\ \frac{2\alpha}{n} & \text{if } n \leq 2\alpha \end{cases}$$

satisfies to $D_V(\eta) = n\eta^2$. In particular, $\eta_V \leq \eta$. The conclusion ensues from some elementary inequalities.

□

Proof of Theorem 5.1. Let us note for all $W_j \in \mathbb{W}_j$, and all integer $\rho_j \in \mathbb{N}^\star$, the set $Z_j(W_j, \rho_j) = W_j \cap \mathcal{B}_{\mathtt{x}}(0, \rho_j)$. Let then, for all $\boldsymbol{W} = (W_1, \dots, W_k) \in \prod_{j=1}^k \mathbb{W}_j$, $\boldsymbol{\rho} = (\rho_1, \dots, \rho_k) \in (\mathbb{N}^\star)^k$, and $K \in \mathcal{K}$ the set

$$V_K(\boldsymbol{W}, \boldsymbol{\rho}) = \left\{ (t,x) \mapsto f_{\pi_K(u(x))}(t),\, u \in \prod_{j=1}^k Z_j(W_j, \rho_j) \right\}.$$

We define

$$\mathbb{V} = \left\{ V_K(\boldsymbol{W}, \boldsymbol{\rho}),\, \boldsymbol{W} \in \prod_{j=1}^k \mathbb{W}_j,\, \boldsymbol{\rho} \in (\mathbb{N}^\star)^k,\, K \in \mathcal{K} \right\}$$

and we define Δ on $\cup_{V \in \mathbb{V}} V$ by

$$\Delta(V_K(\boldsymbol{W}, \boldsymbol{\rho})) = \sum_{j=1}^k \left(\Delta_j(W_j) + \ln\left(2\rho_j^2\right)\right) + \Delta_{\mathcal{K}}(K).$$

We apply Theorem 2.1 with $(\mathbb{V}, \Delta)$ to derive an estimator $\hat{s}$ such that, for all $\boldsymbol{W} = (W_1, \dots, W_k) \in \prod_{j=1}^k \mathbb{W}_j$, $\boldsymbol{\rho} = (\rho_1, \dots, \rho_k) \in (\mathbb{N}^\star)^k$, and $K \in \mathcal{K}$,

$$C\mathbb{E}\left[H^2(s, \hat{s})\right] \leq d_2^2\left(\sqrt{s}, V_K(\boldsymbol{W}, \boldsymbol{\rho})\right) + \eta^2_{V_K(\boldsymbol{W}, \boldsymbol{\rho})} + \frac{\Delta(V_K(\boldsymbol{W}, \boldsymbol{\rho}))}{n}$$

where C is an universal positive constant. It comes from Proposition 6.1 and Lemma 6.8 that

$$C'\eta^2_{V_K(\boldsymbol{W}, \boldsymbol{\rho})} \leq \frac{1}{n} \sum_{j=1}^k \left(\frac{\dim(W_j) \vee 1}{\alpha_j(K)}\right)\left(\ln\left(1 + kR_j(K)\rho_j^{\alpha_j(K)}\right) + \ln n + 1\right)$$

where C' is an universal positive constant. In particular for all $f \in \mathscr{F}$ of the form $f(t,x) = f_{u(x)}(t)$, for all K such that $u(\mathbb{X}) \subset K$, for all map $v = (v_1, \dots, v_k) \in \prod_{j=1}^k W_j$, such that for all $1 \leq j \leq k$, $\|v_j\|_{\mathtt{x}} \leq \|u_j\|_{\mathtt{x}}$, and g of the form $g(t,x) = f_{\pi_K(v(x))}(t)$, the preceding inequality applied with $\rho_j = \inf\{i \in \mathbb{N}^\star,\ i \geq \|u_j\|_{\mathtt{x}}\}$ leads to

$$\begin{aligned} C''\mathbb{E}\left[H^2(s,\hat{s})\right] \leq\ & d_2^2\left(\sqrt{s}, f\right) + d_2^2\left(f,g\right) \\ & + \frac{\sum_{j=1}^k \left(\Delta_j(W_j) + \ln(1+\|u_j\|_{\mathtt{x}})\right) + \Delta_{\mathcal{K}}(K)}{n} \\ & + \frac{1}{n}\sum_{j=1}^k \left(\frac{\dim(W_j)\vee 1}{\alpha_j(K)}\right)\left[\ln\left(1 + kR_j(K)(1+\|u_j\|_{\mathtt{x}})^{\alpha_j(K)}\right) + \ln n + 1\right]. \end{aligned}$$

It ensues from Assumption 5.1 that

$$d_2^2\left(f,g\right) \leq k\sum_{j=1}^k R_j(K)^2 \|u_j - v_j\|_{\mathtt{x}}^{2\alpha_j(K)}.$$

We conclude by choosing v_j the projection of u_j on W_j in the space $\mathbb{L}^2(\mathbb{X}, \nu_n)$. $\square$

Acknowledgement: Many thanks to Yannick Baraud for his suggestions, comments and careful reading of the paper.

References

Aubry, C. and Dabye, A. S. (2001). Asymptotic normality of the minimum distance estimators for a poisson process with a discontinuous intensity function. *Journal of Statistical Planning and Inference*, 99(1):3 – 23.

Baraud, Y. (2010). Estimator selection with respect to hellinger-type risks. *Probability Theory and Related Fields*, pages 1–49.

Baraud, Y. and Birgé, L. (2009). Estimating the intensity of a random measure by histogram type estimators. *Probability Theory and Related Fields*, 143:239–284.

Baraud, Y. and Birgé, L. (2011). Estimating composite functions by model selection. *ArXiv e-prints*.

Birgé, L. (2006). Model selection via testing: an alternative to (penalized) maximum likelihood estimators. *Annales de l'Institut Henri Poincare (B) Probability and Statistics*, 42(3):273 – 325.

Birgé, L. (2007). Model selection for poisson processes. *Lecture Notes-Monograph Series*, 55:pp. 32–64.

Comte, F., Gaïffas, S., and Guilloux, A. (2011). Adaptive estimation of the conditional intensity of marker-dependent counting processes. *Annales de l'Institut Henri Poincaré (B) Probabilités et Statistiques*, 47(4):1171–1196.

Cox, D. R. and Lewis, P. A. (1966). Statistical analysis of series of events.

Dabye, A. S. (1993). *Estimation paramétrique pour un processus de Poisson d'intensité discontinue*. PhD thesis, Université du Maine.

Dahmen, W., DeVore, R., and Scherer, K. (1980). Multi-dimensional spline approximation. *SIAM Journal on Numerical Analysis*, 17(3):pp. 380–402.

Duane, J. T. (1964). Learning curve approach to reliability monitoring. *IEEE Transactions on Aerospace*, 2(2):563–566.

Goel, A. L. and Okumoto, K. (1979). Time-dependent error-detection rate model for software reliability and other performance measures. (3):206–211.

Juditsky, A., Lepski, O., and Tsybakov, A. (2009). Nonparametric estimation of composite functions. *Ann. Stat.*, 37(3):1360–1404.

Lawless, J. F. (1987). Regression methods for poisson process data. *Journal of the American Statistical Association*, 82(399):pp. 808–815.

Massart, P. (2003). *Concentration inequalities and model selection*, volume 1896 of *Lecture Notes in Mathematics*. Springer Berlin/Heidelberg. École d'été de Probabilités de Saint-Flour.

Musa, J. D. and Okumoto, K. (1984). *A logarithmic poisson execution time model for software reliability measurement*. ICSE '84. IEEE Press, Piscataway, NJ, USA.

Reynaud-Bouret, P. (2003). Adaptive estimation of the intensity of inhomogeneous poisson processes via concentration inequalities. *Probability Theory and Related Fields*, 126:103–153.

Yamada, S., Mitsuru, O., and Osaki, S. (1983). S-shaped reliability growth modeling for software error detection. *IEEE Transactions on Reliability*, (5):475–484.

Université de Nice Sophia-Antipolis, Laboratoire J-A Dieudonné, Parc Valrose, 06108 Nice cedex 02, France.

E-mail address: msart@unice.fr